DOI: 10.13382/j.jemi.2017.12.006

基于卷积神经网络的深度图像超分辨率重建方法*

李 伟 张旭东

(合肥工业大学 计算机与信息学院 合肥 230009)

摘 要:为了更有效地提高深度图像的分辨率,构建了一种更深层次的深度图像超分辨率重建的卷积神经网络。该网络直接将低分辨率深度图像作为网络的初始输入,通过卷积神经网络学习图像的高阶表示,获得更具有表达能力的深层特征,同时在网络的输出层引入亚像素卷积层,针对提取到的特征学习不同上采样滤波器,实现上采样放大操作。为了实现网络更好地收敛,在网络中加入了残差网络结构。在 4 个常用数据集上的实验结果表明,与其他先进方法相比,该方法网络收敛速度更快,并可以有效地保护图像的边缘结构,解决伪影问题,且在定性和定量两方面均取得了很好的重建效果。

关键词: 深度图像;超分辨率重建;卷积神经网络;残差网络结构

中图分类号: TP391.41 **文献标识码**: A **国家标准学科分类代码**: 510.4050

Depth image super-resolution reconstruction based on convolution neural network

Li Wei Zhang Xudong

(School of Computer Science and Information, Hefei University of Technology, Hefei 230009, China)

Abstract: In order to improve theresolution of depth imagemore effectively, a deeper convolution neural network is constructed in this paper. The network directly adapts the low-resolution depth image as the initial input of the network, and learns the high-order representation of depth image through the convolution neural network to obtain the features with more expressive ability. At the same time, the sub-pixel convolution layer is introduced at the output layer of the network. Based on the extracted features, a set of sampling filter is learned to achieve the amplification operation. For a better performance of the convergence, the residual network is added to our network. The experimentsare conducted on four commonly used datasets, and the results show that our network is faster than other advanced ones at the convergence rate. The proposed method can effectively protect the edge structure of the depth image, solve the artifact problem, and reachesgreat performance both in qualitative and quantitative aspects.

Keywords: depth image; super-resolution; convolution neural network; residual network

0 引 言

近年来,随着 3D 成像技术的发展,深度信息的获取和处理受到广泛关注。然而,因为深度相机硬件条件自身的限制,其获得的深度图像空间分辨率普遍较低,很难满足工业应用的需求[1-4],例如 SwissRange SR4000 深度相机和 PMD Camcube 相机的分辨率仅只有 200×200,而微软的 Kinect 相机获得的深度图像的分辨率也只有 640×480 远低于相对应的彩色图像分辨率。如何获得高分辨的深度图像,研究人员进行了相关的研究。目前深度图像的超分辨率重建(super resolution,SR)的研究主要集中在多帧深度图像融合的方法[5-7]、同场景彩色图像引导的重建方法[8-11]和基于学习的单幅图像重建方法[12-16]。

多帧深度图像融合的方法是通过融合同场景的多帧

收稿日期:2017-05 Received Date: 2017-05

*基金项目:国家自然科学基金(61403116)、中国博士后基金(2014M560507)资助项目

序列图像,利用多帧图像间互补的相关信息,获得高质量的图像。文献[5]将多幅彩色图像的重建方法应用到深度图像 SR 问题中,通过双边正则化(bilateral regularization)求解深度图像总体能量的最小值,在保持了深度边缘的尖锐性的同时去除了随机噪声的影响。文献[6]遵循贝叶斯框架,提出一种基于深度图像边缘保护的马尔科夫随机场(markov random field,MRF)先验模型,利用边缘自适应的 MRF 先验信息,有效的保护了深度图边缘的不连续性。但该类方法需要通过相机的微小位移来捕获多帧的深度图像,限制了该类方法的实际应用,此外重建效果受相机位姿估计的影响较大。

同场景彩色图像引导的重建方法是当前的研究热点之一。该类方法通过利用预配准的同场景彩色图像中的高频分量(如边缘信息)提供先验信息,协助深度图像的重建。文献[8]最先利用 MRF 模型将同场景彩色图像数据整合到低分辨率深度图像数据中,建立它们的联系,并采用共轭梯度(conjugate gradient,CG)算法求解最小二乘优化问题,有效地提高了深度图像的分辨率。文献[9]利用配准的高分辨率彩色图像作为参考图像,应用双边滤波器(bilateral filter)在迭代优化的过程中来不断提高深度图像的空间分辨率。文献[10]提出在最小二乘的优化框架中结合非局部均值滤波器(nonlocal means filtering, NLM)和边缘加权的方法,有效地保持深度图像的边缘结构。文献[11]利用二阶广义总变分(total generalized variation,TGV)模型,把高分辨率彩色信息作为深度图边缘的约束项进行超分辨率重建,有效地保护了深度图的边缘结构。尽管该类方法重建效果较为理想,但是在实际应用中很难获得与深度图完全配准吻合的高分辨率的彩色图像,限制了彩色图像引导方法的应用。

基于学习的单幅图像重建方法首先通过训练预构造的高/低分辨率图像的数据集,获得高分辨率图像和低分辨率图像之间的对应关系(字典),然后通过字典和低分辨率图像,重建高分辨率图像。文献[12]采用基于图像分块的方法,训练额外的深度图数据集,应用 MRF 模型寻找深度图对应位置的高分辨率图像块。文献[13]通过训练额外的高低分辨率深度图像数据集,利用稀疏编码的方法学习得到具有边缘先验信息的字典,然后再利用 TGV 模型优化重建高分辨率深度图。文献[14]在 MRF 模型中加入低分辨率深度边缘约束信息优化重建,有效地保持深度图像的边缘。但该类方法一般都需要进行预处理操作如块的提取和整合,而且难以建立高/低分辨率图像块正确的映射关系,重建结果会出现模糊和伪影的问题。

近年来,深度学习受到了广泛的关注,卷积神经网络(convolutional neural network,CNN)作为深度学习模型框架的一种,由于其出色的学习能力,能够从海量训练数据中自动学习目标特征,已经被成功的运用到图像的超分辨率重建任务。文献[17]首次提出一种用于单幅图像超分辨率重建的深度学习方法,该方法 CNN 框架以端对端方式直接学习高/低分辨率图像之间的映射关系。其网络模型命名为 SRCNN,共由 3 层卷积层组成,分别是特征提取、非线性映射、高分辨率图像重建。但该网络学习到的特征少且单一,同时网络收敛速度慢。而文献[18]将 SRCNN 作为网络的基础单元,提出一个递进的深层卷积神经网络结构可以不断学习深度图像的高频信息(如边缘信息),然后利用深度域统计和彩色图引导的方法优化网络学习得到的深度图像。但其网络结构并不是严格意义上的端对端的网络,需要将前一个 SRCNN 单元输出结果再输入到下一个 SRCNN 单元,中间还需要额外的预处理过程,并且通过级联重建效果提升不明显,还增加了网络的训练时间。但值得一提的是深度学习的方法普遍优于现有的传统重建方法。

综上所述,考虑到重建性能和网络训练时间两方面,本文提出了一种深度学习的深度图像超分辨率重建方法。该方法构建了一个新的端对端的卷积神经网络模型。该网络模型由卷积层、ReLU 层[19]、残差网络结构[20]和亚像素卷积层[21]组成,以低分辨率深度图像作为输入,高分辨率深度图像作为输出,构成了由低分辨率图像到原始图像的非线性映射,实现了深度图像的重建。实验结果表明本文方法在提升了重建性能的同时缩短了网络的训练时间。

1　基于 CNN 的深度图像的超分辨率重建

深度图像(depth image)又称为距离图像(range image),其图像上的每个像素值表示景深的远近。而深度图像的超分辨率重建是指由单幅或多幅低分辨率图像重建出分辨率更高、携带信息细节更丰富的图像的过程,其降质模型可以用式(1)来表示。

$$Y = \boldsymbol{D}\boldsymbol{H}x + n \tag{1}$$

式中:x 表示高分辨率深度图像,Y 表示低分辨率深度图像,$\boldsymbol{D}$ 和 $\boldsymbol{H}$ 分别表示降采样因子和模糊矩阵,n 表示加性噪声,主要包括系统噪声、随机噪声。

深度图像的超分辨率重建也可以看成是将低分辨率深度图像映射到高分辨率深度图像的过程。换句话说,深度图像的超分辨率重建也就是为映射的拟合过程。而 CNN 可以通过网络训练过程不断缩小网络输出和高分辨率深度图像的偏差,具有较强的映射逼近能力。CNN 是一种采用多层层次结构网络的具有鲁棒性的深度学习模型,其采用的局部连接和权值共享的方式,减少了权值数量使得网络易于优化。而 CNN 层次之间的紧密联系

和空间信息使得其特别适用于图像的处理。因此,CNN 成为研究图像的理想模型,并取得显著的效果[22-26]。

图 1 所示为典型的基于 CNN 的图像超分辨率重建的框架[17]。但该网络对于深度图像的超分辨率重建任务仍存在局限性。原因如下:该网络模型首先对低分辨率图像进行双三次插值(Bicubic)操作,将其放大到和高分辨率图像同样的尺寸大小。而双三次插值的方法是在假定图像平滑的前提下,预测未知的高分辨率像素点,会破坏原始图像中的高频细节信息。并且深度图像与普通彩色图像不同,其图像上的像素值所表示的是景深的远近(距离信息),而不是彩色图像提供的彩色信息。所以对深度图像进行插值获得的像素信息没有实际的深度意义。就深度图像自身而言,其包含的纹理信息较少、且在深度不连续区域(如边缘)变化更加尖锐。这样的预处理操作,会造成重建的深度图整体模糊,尤其在深度不连续区域会出现细节的丢失。由图 1 可知,该网络结构较为简单,能够学习到的特征少且单一,导致重建性能较差。

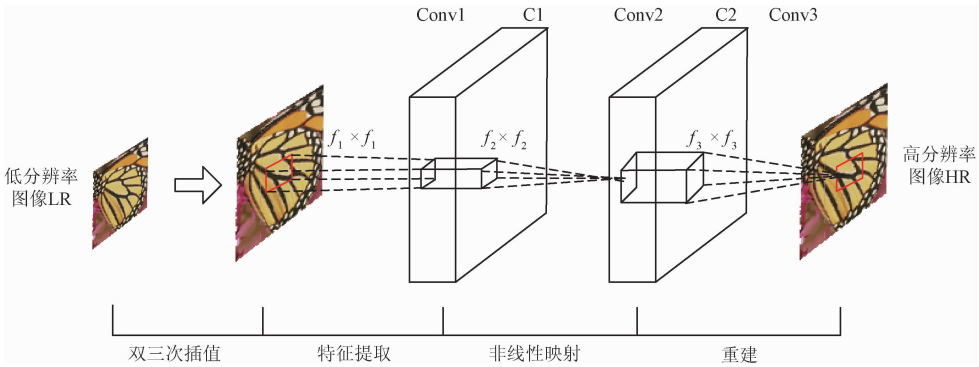


图 1　典型卷积神经网络的框架

Fig. 1　The frameworkof typical convolution neural network

为了解决这些问题,本文直接将低分辨率图像作为网络的初始输入,利用 CNN 充分提取深度图的原始信息。同时引入亚像素卷积层,针对提取的不同特征学习一组上采样滤波器,代替手工设计的插值操作。同时为了更有效地学习深度图像更多的微小结构和细节信息,进一步提升重建性能。本文构建了一个更深层次的网络结构,但加深网络层数伴随而来的是网络收敛难的问题。因此,在网络中加入残差网络结构,帮助整个网络更快的收敛。

1.1　网络结构

深度图像超分辨率重建的 CNN 设计应充分考虑低分辨率深度图像和高分辨率深度图像的关系,即构建的网络实际上是一种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。图 2 所示为本文构建的一种深度图像重建的卷积神经网络。该网络共有 5 层,主要由特征提取,非线性映射和亚像素卷积重建 3 部分组成。其中,网络的隐含层均由特征图组成。

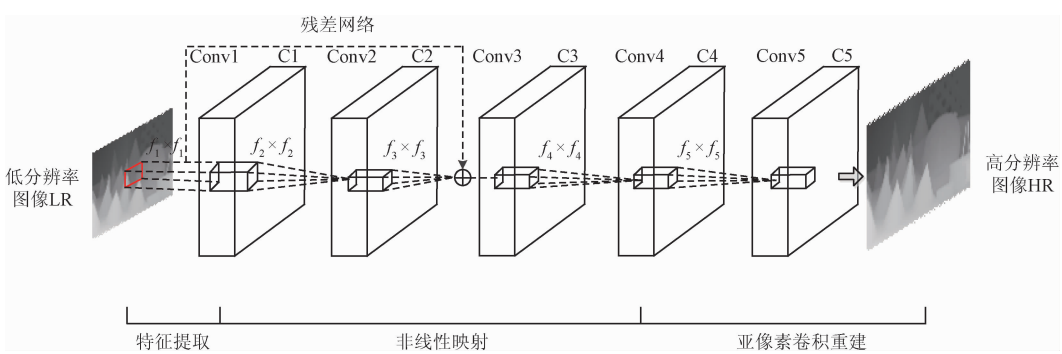


图 2　本文构建的卷积神经网络框架

Fig. 2　The frameworkof theconvolution neural network

卷积层(Conv)1 为特征提取层。低分辨率深度图像输入网络后,使用 64 个尺寸大小为 5 ×5 的卷积核对原始的低分辨图像进行卷积操作,提取深层特征,形成 C1 层,包含 64 幅特征图。随后激活函数(ReLU)对提取到的特征进行非线性处理(用来保证网络更快的收敛),具体操作如下:

$$F_1(Y) = \phi(W_1 * Y + b_1) \tag{2}$$

式中：$*$ 表示卷积操作，W_1、b_1 分别表示 Conv1 的卷积核和偏置向量，W_1 大小为 $c \times n_1 \times f_1 \times f_1$，其中 n_1 是该层的滤波器数量，$c=1$（处理的深度图像为一般为灰度图像，单通道），$f_1 \times f_1$ 是卷积核的大小，ϕ 代表激活函数，Y 表示原始的低分辨率图像，$F_1(Y)$ 表示该层输出的特征图，即 C1 层。

Conv2、Conv3、Conv4 为非线性映射层。C1 层的特征图作为残差网络的输入，然后将残差网络的输出 C3 层的特征图与 32 个大小为 3×3 的卷积核进行卷积操作，形成包含 32 幅特征图的 C4 层，建立了 C1 ~ C4 层的非线性映射关系。其中，Conv2、Conv3 的公式和 Conv1 相同，Conv4 的公式如下：

$$F_4(Y) = \phi(W_4 * (F_1(Y) + F_3(Y)) + b_4) \tag{3}$$

式中：W_4、b_4 分别表示 Conv4 的卷积核和偏置向量，W_4 大小为 $n_3 \times n_4 \times f_4 \times f_4$，$F_3(Y)$ 表示 Conv3 输出的特征图，即 C3 层。

Conv5 为亚像素卷积层。C4 层通过亚像素卷积大小为 3×3 的卷积核，生成 r^2 幅特征图（上采样倍数 r），形成 C5 层，并通过像素的重新排列生成重建后的高分辨率深度图像，具体操作如下：

$$I^{SR} = F_5(Y) = SF(W_5 * F_4(Y) + b_5) \tag{4}$$

式中：W_5、b_5 分别表示 Conv5 的卷积核和偏置向量，W_5 大小为 $n_4 \times n_5 \times f_5 \times f_5$，$F_4(Y)$ 表示 Conv4 输出的特征图（C4 层），SF 表示像素的重新排列操作，$F_5(Y)$ 表示重建后的高分辨率深度图像。

该网络具有如下特点：

1）输入低分辨率图像，提取图像的原始信息。典型网络的先插值操作并不是最优的上采样方法，会影响低分辨率深度图像的原始信息。因此本文将低分辨率图像作为网络的初始输入，使得后续的卷积操作都是在高分辨率空间下进行，进一步降低了计算复杂度。

2）加入残差网络结构，加快整个网络的收敛过程。由于加深网络层数，会带来收敛难的问题，本文在非线性映射过程中加入残差网络结构以用来更容易的学习高/低分辨率图像之间的映射关系，加快整个网络的收敛过程。

3）引入亚像素卷积，实现上采样放大操作。通过亚像素卷积层学习一组上采样滤波器，更加有效地针对提取到的不同的特征图进行上采样放大，输出多幅高分辨率特征图，最后通过像素的重新排列生成高分辨率深度图像。

1.2　残差网络结构

非线性映射是指将提取到的特征图映射到其他高维特征图的过程，通过加深非线性映射过程中的网络层数可以实现更好重建效果，但随着隐含层数的增加，反向传播传递给较低层的信息会越来越少，即会出现梯度爆炸/梯度消失和网络性能退化的问题[20]。因此，本文在非线性映射过程中加入残差网络结构（residual network，resnet），更加有效的实现梯度的下降，加快整个网络的收敛过程，并且整个网络未引入额外的参数和计算复杂度。

典型的 N 层残差网络结构如图 3 所示，本文使用的残差网络结构只由卷积层，激活函数，快捷连接（skip connections）3 部分组成，舍弃了批量化（batch normalization，BN）单元。考虑到结构复杂度，训练时间等因素，采用与文献[27]相似的小尺寸的卷积核。假设是残差结构的输入，是残差结构的输出，残差网络结构可用下式描述，其中是残差网络结构的参数，是网络学习到的残差映射关系：

$$y = x + f(\theta, x) \tag{5}$$

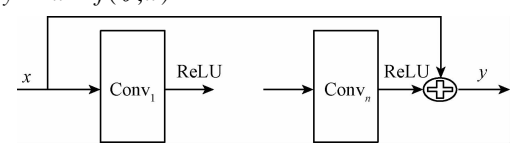


图 3　典型的 N 层残差网络结构

Fig. 3　Thestructure of a typical N-layer residual network

1.3　亚像素卷积

亚像素卷积重建是将非线性映射得到的高维特征图上采样重建成高分辨率深度图像的过程。上采样操作承接非线性映射和重建两个部分，在整个重建过程中起到关键的作用。

图 4 所示为 HR 网络和 LR 网络的卷积重建过程：这里把先使用插值放大操作的网络（如 SRCNN 网络）称为高分辨率 HR 网络，本文的网络称为低分辨率 LR 网络。假设低分辨率图像的大小是 2×2，放大系数为 2。图 4（a）为 HR 网络先对原始低分辨图像进行插值放大，得到大小为 5×5 图像（为了简化示意，只取了两边填充，所以图示为 4×4，其中灰色小块表示插值操作），随后图像和大小为 2×2 卷积核进行卷积操作，得到 4×4 的高分辨率图像。图 4（b）为 LR 网络直接将原始图像块和卷积核

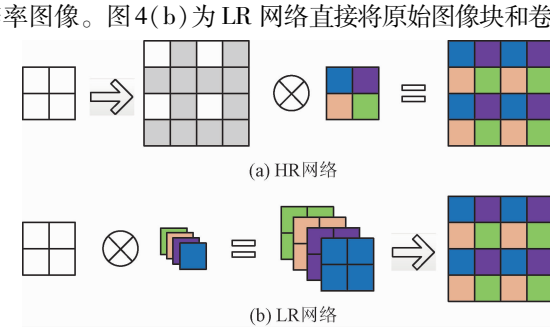


图 4　HR 网络和 LR 网络的卷积重建过程

Fig. 4　The convolution reconstruction processof HR and LR network

大小为 1×1 进行卷积操作,得到 4 个大小为 2×2 图像块,然后通过像素的重新排列生成和 HR 网络相同的高分辨率图像。

假设亚像素卷积重建之前每层有 n 幅特征图,那么 SRCNN 应有 n/r^2 幅特征图。两种网络分析如表 1 所示,其中 W、H 为图像高,宽 l 是网络层数,其中放大系数是 r,卷积核尺寸是 $f\times f$,可以看出两种网络具有相同网络复杂度,但本文网络参数更多,可以学习到更多的深层特征。因此在相同的训练时间内本文网络重建性能要优于 SRCNN。

表 1　HR 网络和 LR 网络的网络分析

Table 1　The analysis of HR and LR network

	网络复杂度	网络参数
HR 网络	$\frac{n}{r^2}\times\frac{n}{r^2}\times f_r\times f_r\times W\times H$	$l\times\frac{n}{r^2}\times\frac{n}{r^2}\times f_r\times f_r$
LR 网络	$n\times n\times f\times f\times\frac{W}{r}\times\frac{H}{r}$	$l\times n\times n\times f\times f$

1.4　网络训练

本文基于反向传播的随机梯度下降法[28]来训练网络,并采最小均方差(mean squared error,MSE)作为损耗函数。为使网络收敛,令下式取最小值:

$$L(\Theta)=\frac{1}{N}\sum_{i=1}^{N}\|F(Y_i;\Theta)-X_i\|^2 \tag{6}$$

式中:$L(\Theta)$表示最小均方差,$\Theta=\{W_i,b_i\}$ 表示要训练学习的网络参数,$F(Y_i;\Theta)$表示重建深度图像,X_i 表示参考图像。

2　实验结果及分析

本节从定性和定量两方面进行了 2 组对比实验。1)在 Middlebury[29]数据集上,对有无残差网络结构进行对比实验,分析残差网络结构对重建性能的影响。2)为了验证本文方法的有效性,在 4 种常见的公共数据集 Middlebury[29-32]、Sysnthesis[12]、RGBZ[33]和 RGBD[34]数据集上与已有的深度图像超分辨率重建方法进行比较。在此基础上,分析不同迭代次数对重建结果的影响。

2.1　实验设置

本文实验的 PC 配置为 NVIDIA GeForce TITANX、RAM 16 G、Linux 64 位操作系统,编译软件为 MATLAB 2014a。

网络参数:每层卷积层滤波器尺寸除 Conv1 为 5×5,其余的滤波器尺寸都为 3×3,滤波器数量 $n_1=n_2=n_3=64$,$n_4=32$,$n_5=r^2$。本文利用 Caffe[35]搭建卷积神经网络,每层网络参数以高斯分布进行初始化,其均值为 0,标准差为 10^{-3},虽选用较大的学习率可以使网络尽快收敛,但可能出现局部最优的问题,因此与 SRCNN 相同,本文的网络初始学习率设置为 10^{-4},最后一层学习率设置为 10^{-5},且最大迭代次数都设置为 8×10^6。

训练数据集:本文选用的训练数据集由 4 部分组成,分别是 Middlebury、Sysnthesis、RGBZ 和 RGBD 数据集。Middlebury 数据集包含丰富的高分辨率深度信息,图像边缘清晰,是目前验证超分辨率重建算法优劣常用的数据集,从中随机抽取 34 幅不同场景的深度图像;Sysnthesis 数据集是合成的 3D 场景的深度图数据集,其特点是具有较好的边缘特征,仅含有少量的噪声,同样随机抽取 23 幅深度图像;RGBZ 和 RGBD 数据集是 TOF 拍摄的真实场景深度图像,从中抽取了 15 幅深度图像。为了充分利用这些深度图像,对数据集进行增强(旋转 90°),得到 144 幅边缘清晰的高分辨率深度图像。同时避免直接使用大尺寸深度图像,先对深度图像进行分块,这样的数据处理不仅不会影响网络的重建性能,还会减少网络的训练时间[36]。对所有的上采样放大倍数(2、3、4)本文都采用相同的分块步长 14,最终得到 111 360 个大小为 25×25 的子图像块,作为网络的初始输入。

评价指标:为了有效地评价重建效果的优劣,本文采用均方根误差(root mean squared error,RMSE)和结构相似性(structural similarity index measurement,SSIM)作为客观评价指标。均方根误差 $RMSE$,即均方误差的开平方,是图像质量客观评价标准的一种,首先通过重建结果与标准参考图像进行均方误差运算,最后进行平方根,其值越小,表明重建结果图像质量越好。而结构相似性是指两副图像之间的结构信息相似程度,$SSIM$ 值越接近 1,说明重建的深度图与原图越相似,重建效果越好。

$$RMSE=\sqrt{\frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}(Y(i,j)-X(i,j))^2} \tag{7}$$

$$SSIM=\frac{(2\mu_X\mu_Y+C_1)(\sigma_{XY}+C_2)}{(\mu_X^2+\mu_Y^2+C_1)(\sigma_X^2+\sigma_Y^2+C_1)} \tag{8}$$

式(7)中图像的大小 $M\times N$,是重建后的高分辨率图像,X 是参考图像;式(8)中 μ_X 和 μ_Y 分别为参考图像的灰度平均值和方差;σ_X 和 σ_Y 分别是重建后的高分辨率图像的灰度平均值和方差;σ_{XY}为参考图像和重建后的高分辨率图像的协方差;C_1 和 C_2 为常数,防止分母为 0。

2.2　残差网络结构对重建性能的影响

对有无残差网络结构对重建性能的影响进行了实验分析,在网络中包含残差网络结构的实验部分中采用了两种常用的残差网络结构,分别是 2 层残差网络结构:3×3,3×3(用 3-3 表示)和 3 层残差网络结构:1×1,3×3,1×1(用 1-3-1 表示),其中所有卷积层滤波器数量都设置为 64。

图 5 所示为有无残差网络结构在 Middlebury 数据集

上的重建结果。3-3(no residual)表示未加入残差网络。从视觉观测,图5(b)中双三次插值(Bicubic)的重建结果边缘模糊不清,效果最差;1-3-1(no residual)网络训练难以收敛,重建的效果不如双三次插值,这里不参与比较,但当加入残差网络结构后,网络可以实现收敛并提高了重建性能,结果如图 5(c)所示;从图 5(d)3-3(no residual)和图5(e)3-3 的对比,可以看出在网络加入残差结构可以获得更好的重建效果,图像边缘更加清晰,重建结果接近真值;图5(f)3-3-3-3 是两个 3-3 残差网络结构的级联的重建结果,相比 3-3 得到小幅提升。

图 5　有无残差网络结构在 Middlebury 数据集上的重建结果(放大倍数为 3)

Fig. 5　Thereconstruction results with/without the residual network on the Middlebury dataset at a magnification of 3

表 2 给出有无残差网络结构在 Middlebury 数据集上重建结果的定量分析,上采样放大倍数为 3,其中黑色字体表示最优值,下划线标识的是次优值。由结果可见,在网络中加入残差网络结构有助于重建性能的提升,如 3-3 与 3-3(no residual)相比,RMSE 平均降低约 0.12,SSIM 指标评价提升约 0.003,并且 3-3 的残差网络结构重建效果优于 1-3-1 残差网络结构。随着 3-3 残差结构单元数量增加(表 2 中的 3-3-3-3),重建效果也能得到小幅提升,但是因为网络层数和参数的增加,网络所需要的训练时间变长。

表 2　有无残差网络结构在 Middlebury 数据集上重建结果的定量分析

Table 2　Quantitative analysis of the reconstruction results with/without the residual network on the Middlebury dataset

	RMSE			SSIM		
	cones	teddy	venus	cones	teddy	venus
Nearest	4.627 8	3.639 7	2.382 6	0.941 3	0.948 9	0.981 2
Bicubic	3.165 1	2.441 1	1.662 5	0.960 5	0.965 9	0..984 5
1-3-1	2.383 0	1.867 5	0.991 1	0.971 2	0.975 6	0.993 5
3-3(no residual)	2.179 0	1.721 5	0.851 3	0.976 9	0.979 9	0.994 9
3-3	2.051 1	1.676 8	0.769 6	0.979 9	0.981 5	0.996 4
3-3-3-3	**1.983 1**	**1.603 2**	**0.724 0**	**0.980 3**	**0.984 6**	**0.996 9**

本实验验证了残差网络结构不仅有助于重建性能的提升,还可以加快网络的收敛。综合训练时间和重建性能的考虑,采用 3-3 作为本文的残差网络结构。

2.3　与其他方法的比较

为了验证本文算法的有效性,本文分别在 Middlebury、Laser Scan 数据集和真实场景 RGBZ、RGBD 数据集上进行实验分析。

实验 1 在 Middlebury 和 Laser Scan 数据集常用的测试深度图像与最邻近插值(Nearest)、双三次插值(Bicubic)和 Aodha[12]、Ferstl[13]、Xie[14]等具有代表性的算法以及最近提出的稀疏字典学习方法 Yang[22]、深度学习方法 SRCNN[17],Song[18]进行比较(因文献[18]作者未提供源码,只引用文献中给出的定量的实验结果)。

图 6、7 所示为本文方法和上述其他方法在上采样放大倍数为 4 时的重建结果。由于 Aodha[12]需要对重建后的高分辨率图像块进行拼接,因此尤其在深度不连续的区域常会出现错误拼接,而 Xie[14]将低分辨率深度图像边缘作为约束条件,难以获得边缘的细节特征,图 6(b)、(c)所示的是这两种方法重建的结果,可以看出在深度不连续区域都出现明显的边缘失真问题。

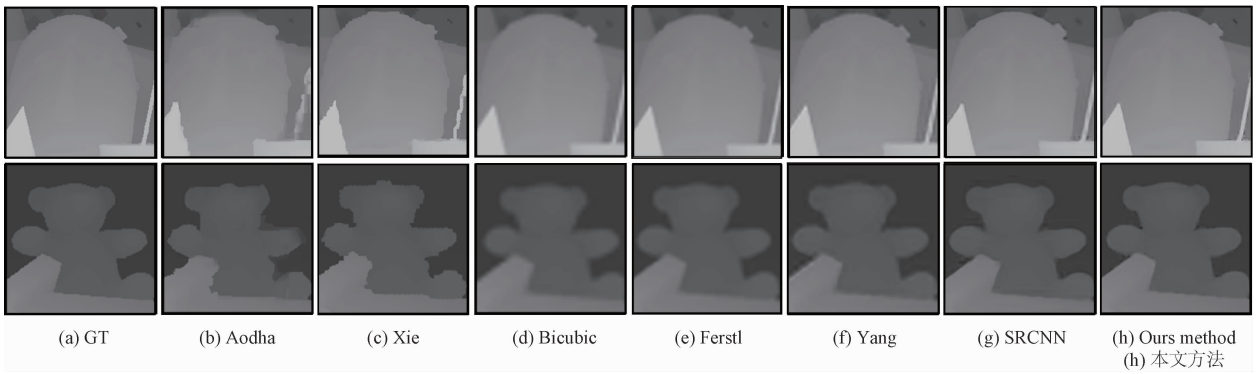


图 6　不同方法在 Middlebury 数据集上的重建结果(放大倍数为 4)

Fig. 6　The reconstruction results with the different methods on the Middlebury dataset at a magnification of 4

图 7　不同方法在 Laser Scan 数据集上的重建结果(放大倍数为 4)

Fig. 7　The reconstruction resultswith the different methods on the Laser Scan dataset at a magnification of 4

Ferstl[13]采用滤波的方法进行快速重建,但由于其滤波器是高斯核函数组成,重建结果会过度平滑问题,其是在深度不连续区域会出现边缘模糊的情况,而双三次插值是基于平滑假设的原理,同样也出现了边缘模糊的情况。Yang[22]和 SRCNN[17]重建性能较好,但由于 Yang[22]字典大小和 SRCNN[17]网络参数的限制,学习得到的特征少且较为单一,因此重建结果细节较差,由图 7(d)、(e)可以看到图像边缘出现了不同程度的伪影问题;而本文方法通过更深的网络结构,学习到更多的高频特征信息,重建的深度图视觉效果更好,接近真值,在避免边缘模糊的同时有效地解决边缘失真和伪影的问题。

表 3、4 给出了不同方法在上采样倍数为 2、4 时的重建结果的定量分析,表 5 给出了不同方法在上采样倍数为 4 时的重建结果的定量分析。其中黑色字体表示最优值,下划线标识的是次优值。由结果可见,本文方法在两个客观指标上都取得了最优值,与定性评价结果一致。Song[18]和本文方法都是基于 CNN 对深度图像的超分辨率重建,但 Song[18]需要对低分辨图像先进行插值放大,这样会影响图像的原始信息,和其相比,本文的 RMSE 指数平均降低了约 0.21,SSIM 指数平均提升约 0.003。

为了进一步验证本文方法的有效性,实验 2 在 RGBZ 数据集和 RGBD 数据集与 5 种具有代表性的图像超分辨率重建方法进行对比。不同方法在上采样放大倍数为 4 时的重建结果如图 8 所示,由结果可见,本文效果优于相比较的几种方法。在避免 Xie[14]重建结果的失真现象的同时,有效地去除了双三次插值的边缘模糊和 Yang[22]、SRCNN[17]的伪影问题。表 6、7 所示为不同方法在上采样放大倍数为 4 时的重建结果的定量分析,其中同样用黑色字体表示最优值,下划线标识的是次优值,可以发现本文方法同样在两个客观指标上都取得了最优值。

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

Table 3　Quantitative analysis of the reconstruction results with the different methods on the Middlebury dataset (RMSE)

	Cones		Teddy		Venus	
	X2	X4	X2	X4	X2	X4
Nearest	4.130 6	5.918 0	3.213 8	4.477 2	2.112 2	2.882 4
Aodha[12]	4.262 2	5.739 9	3.233 2	4.030 4	2.175 3	2.597 3
Xie[14]	3.264 4	5.020 5	2.505 5	4.043 4	1.454 0	2.502 7
Bicubic	2.526 3	3.846 7	1.957 6	2.861 1	1.339 3	1.939 7
Ferstl[13]	2.185 0	3.497 7	1.694 1	2.595 3	1.107 7	1.741 9
Yang[22]	2.095 9	3.440 9	1.609 0	2.546 6	1.035 4	1.627 4
SRCNN[17]	1.658 1	3.325 3	1.313 3	2.108 9	0.582 1	1.003 4
Song[18]	1.435 6	2.978 9	1.197 4	1.800 6	0.559 2	0.879 6
本文方法	**1.204 1**	**2.623 7**	**0.981 8**	**1.745 2**	**0.428 8**	**0.637 2**

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

Table 4　Quantitative analysis of the reconstruction results with the different methods on the Middlebury dataset (SSIM)

	Cones		Teddy		Venus	
	X2	X4	X2	X4	X2	X4
Nearest	0.954 4	0.918 1	0.960 1	0.928 0	0.985 4	0.973 9
Aodha[12]	0.947 2	0.917 2	0.957 7	0.931 4	0.982 7	0.975 6
Xie[14]	0.969 6	0.937 3	0.973 0	0.938 5	0.992 2	0.980 3
Bicubic	0.9745	0.944 2	0.977 9	0.952 4	0.989 8	0.978 7
Ferstl[13]	0.980 2	0.952 6	0.982 9	0.958 3	0.992 3	0.981 9
Yang[22]	0.982 0	0.954 1	0.984 1	0.961 9	0.995 1	0.986 0
SRCNN[17]	0.986 5	0.957 6	0.987 1	0.964 4	0.998 0	0.992 2
Song[18]	0.988 1	0.959 4	0.989 2	0.979 3	0.998 2	0.995 3
本文方法	**0.991 3**	**0.966 2**	**0.991 9**	**0.981 2**	**0.998 5**	**0.997 0**

表 5　不同方法在 Laser Scan 数据集上重建结果的定量分析

Table 5　Quantitative analysis of the reconstruction results with the different methods on the Laser Scan dataset

	RMSE			SSIM		
	Scan21	Scan30	Scan42	Scan21	Scan30	Scan42
Nearest	10.386 6	8.185 7	10.526 0	0.947 1	0.951 1	0.949 8
Xie[14]	9.193 5	7.414 8	8.909 3	0.949 8	0.952 1	0.950 3
Bicubic	6.474 3	5.208 5	6.402 8	0.954 4	0.961 1	0.952 7
Yang[22]	5.363 2	4.443 7	5.277 2	0.950 3	0.959 2	0.949 6
SRCNN[17]	4.682 3	3.756 9	2.518 2	0.958 8	0.966 8	0.968 2
本文方法	**3.179 4**	**2.850 0**	**2.856 3**	**0.967 2**	**0.976 5**	**0.980 6**

(a) GT　(b) Bicubic　(c) Xie　(d) Yang　(e) SRCNN　(f) Our method (f) 本文方法

图 8　不同方法在 RGBZ 数据集和 RGBD 数据集上的重建结果(放大倍数为 4)

Fig. 8　The reconstruction resultswith the different methods on the RGBD and RGBZ dataset at a magnification of 4

表 6　不同方法在 RGBZ 数据集上重建结果的定量分析

Table 6　Quantitative analysis of the reconstruction results with the different methods on the RGBZ dataset

	RMSE			SSIM		
	01	05	09	01	05	09
Nearest	13.890 6	11.442 0	14.609 6	0.900 4	0.944 6	0.926 8
Xie[14]	12.796 3	10.614 2	14.932 0	0.914 5	0.954 8	0.928 6
Bicubic	8.805 2	7.201 0	8.657 6	0.916 1	0.945 7	0.940 8
yang[22]	7.399 4	6.092 4	7.015 5	0.919 5	0.941 6	0.934 5
SRCNN[17]	5.378 7	3.839 6	6.274 2	0.945 9	0.964 1	0.942 6
本文方法	**4.294 0**	**3.353 1**	**4.149 6**	**0.962 4**	**0.978 7**	**0.996 0**

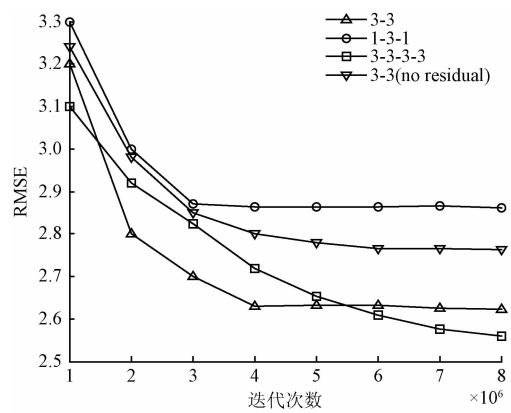
表 7　不同方法在 RGBD 数据集上重建结果的定量分析

Table 7　Quantitative analysis of the reconstruction results with the different methods on the RGBD dataset

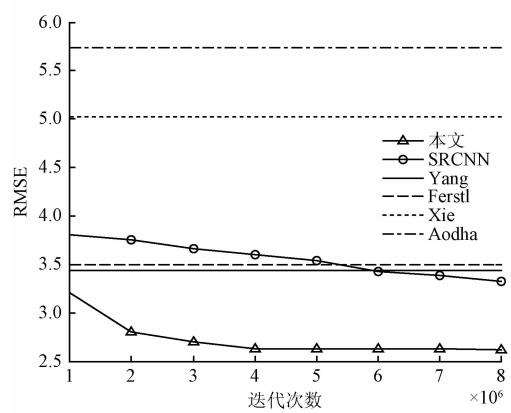
	RMSE			SSIM		
	01	02	06	01	02	06
Nearest	3.706 2	6.245 3	4.305 5	0.957 6	0.956 3	0.971 2
Xie[14]	3.680 9	5.923 4	3.895 6	0.971 0	0.966 8	0.979 8
Bicubic	2.215 0	3.575 0	2.355 0	0.979 6	0.974 5	0.981 2
Yang[22]	1.733 7	2.685 0	1.761 9	0.981 6	0.975 2	0.981 5
SRCNN[17]	1.741 8	3.013 4	1.593 3	0.984 3	0.975 8	0.987 8
本文方法	**1.391 9**	**2.097 6**	**1.145 4**	**0.985 6**	**0.979 0**	**0.988 6**

为了更加直观地分析不同迭代次数对重建结果的影响，给出了有无残差网络结构在 Middlebury 数据集上的 RMSE 随迭代次数的变化趋势，如图 9(a)所示。表 8 又给出了不同网络的训练耗时对比，表中最短的训练时间用黑体表示，次短的训练时间用下划线标识。结合图 9(a)和表 8，可以看出 1-3-1 收敛速度最快且用时最短，3-3 次之，但 3-3 的 RMSE 值远小于 1-3-1；而表中 3-3-3-3 的 RMSE 值最小，但训练时间过长，迭代到 8×10^6 才接近收敛；通过比较 3-3 和 3-3(no residual)两条曲线，可以看出加入残差网络之后可以加快收敛速度，提升重建效果，进一步验证了残差网络结构的优势。

图 9(b)给出了不同方法在 Middlebury 数据集上的 RMSE 随迭代次数的变化趋势。从图中可以看出，传统方法不考虑迭代次数这个因素的影响，所以它们的 RMSE 保持不变，在图上呈直线表示，SRCNN 最终的重建结果优于其他几种算法，而本文方法相对而言是最优的。根据网络复杂度的计算公式[37] $O(\sum_{l=1}^{L} n_{l-1} f_l^2 n_l)$，得到本文网络的复杂度为 96 532，SRCNN 的网络复杂度为 8 032，因此本文网络每次迭代的计算量较大，即单次迭代耗时较长。但由图 9(b)和表 8 可以看出 SRCNN 在迭代次数为 8×10^6 才接近收敛，总用时为 17.11 h，而本文网络在迭代次数为 4×10^6 就已经收敛，且总用时更短(15.03 h)。表明本文的方法可以在较短的训练时间，获得更好超分辨率重建效果。



(a) 有无残差网络结构的重建结果对比曲线
(a) The comparison of the reconstructed results with/without the residual network



(b) 不同方法的重建结果对比曲线
(b) The comparison of the results with the different methods

图 9　迭代次数对实验结果的影响

Fig. 9　The influence of iterations on the experimental results

表 8　不同网络的训练耗时对比

Table 8　Training time-consuming comparison of different networks

	1×10^3/ s	3×10^6/ h	4×10^6/ h	6×10^6/ h	8×10^6/ h
SRCNN[17]	7.97	6.64	8.86	13.28	17.11
3-3-3-3	20.88	17.40	23.20	34.80	46.40
3-3(no residual)	12.53	10.44	13.92	20.88	–
3-3	13.53	11.28	15.03	–	–
1-3-1	12.67	10.56	–	–	–

3　结　论

为了有效提高深度图像的分辨率,同时获得更好的重建效果。本文提出一种基于 CNN 的深度图像的超分辨率重建方法。为了获得更多的细节信息,本文的方法不同于一般 CNN 在高分辨率空间提取特征,而是直接将低分辨率深度图像直接输入到网络中,在低分辨率空间下提取特征。同时引入亚像素卷积层来学习一组上采样滤波器,实现上采样放大操作。考虑到加深网络层数所带来的收敛难的问题,加入了残差网络结构帮助网络更快的收敛。实验结果表明在重建效果和客观指标两方面,本文提出的方法较其他算法都有所提高。且与 SRCNN 相比,在较少的训练时间的情况下达到更优的结果,提高了算法的效率。考虑到同场景的彩色图像能够提供其他的先验信息,因此下一步工作将研究如何将彩色图像加入到网络中,如双通道网络,进一步提高重建性能。

参考文献

[1] CUI Y, SCHUON S, CHAN D, et al. 3D shape scanning with a time-of-flight camera[C]. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2010: 1173-1180.

[2] 邸维巍,张旭东,胡良梅,等. 彩色图约束的二阶广义总变分深度图超分辨率重建[J]. 中国图象图形学报, 2014,19(8):1162-1167.
DI W W, ZHANG X D, HU L M, et al. Depth image super-resolution based on second-order total generalized variation constrained by color image[J]. Journal of Image and Graphics, 2014, 19(8): 1162-1167.

[3] 严徐乐,安平,郑帅, 等. 基于边缘增强的深度图超分辨率重建[J]. 光电子·激光,2016(4):437-447.
YAN X L, AN P, ZHENG SH, et al. Super-resolution reconstruction for depth map based on edge enhancement[J]. Journal of Optoelectronics · Laser, 2016(4): 437-447.

[4] 杨宇翔,汪增福. 基于彩色图像局部结构特征的深度图超分辨率算法[J]. 模式识别与人工智能, 2013, 26(5):454-459.
YANG Y X, WANG Z F. Depth map super-resolution based on the local structural features of color image[J]. Pattern Recognition & Artificial Intelligence, 2013, 26(5):454-459.

[5] SCHUON S, THEOBALT C, DAVIS J, et al. High-quality scanning using time-of-flight depth superresolution[C]. IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, 2008: 1-7.

[6] RAJAGOPANLAN A N, BHAVSAR A, WALLHOFF F, et al. Resolution enhancement of pmd range maps[C]. Joint Pattern Recognition Symposium, 2008: 304-313.

[7] 殷明,庞纪勇,褚标,等. 基于插值与 NSQCT 域融合的图像超分辨率重建[J]. 电子测量与仪器学报,2015, 29(10):1440-1448.
YIN M, PANG J Y, CHU B, et al. Image super-resolution reconstruction based on interpolation and NSQCT fusion[J]. Journal of Electronic Measurementand Instrumentation, 2015, 29(10): 1440-1448.

[8] DIEBEL J, THRUN S. An application of markov random fields to range sensing[C]. Advances in Neural Information Processing Systems, 2005: 291-298.

[9] YANG Q, YANG R, DAVIS J, et al. Spatial-depth super resolution for range images[C]. IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 2007: 1-8.

[10] PARK J, KIM H, TAI Y W, et al. High quality depth map upsampling for 3D-TOF cameras[C]. IEEE International Conference on Computer Vision (ICCV), 2011: 1623-1630.

[11] 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.

[12] MAC AODHA O, CAMPBELL N, NAIR A, et al. Patch based synthesis for single depth image super-resolution[C]. Computer Vision-ECCV, 2012: 71-84.

[13] FERSTL D, RUTHER M, BISCHOF H. Variational depth superresolution using example-based edge representations[C]. Proceedings of the IEEE International Conference on Computer Vision, 2015: 513-521.

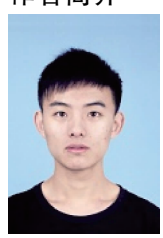
[14] XIE J, FERIS R S, SUN M T. Edge-guided single depth image super resolution[J]. IEEE Transactions on Image Processing, 2016, 25(1): 428-438.

[15] 陶永耀,赵新中,梁婉文,等. 基于 MCA 分解和样本聚类的超分辨率算法[J]. 电子测量技术,2016,39(6): 57-60.
TAO Y Y, ZHAO X ZH, LIANG W W, et al. The SR algorithm based on MCA decomposition and sample clustering[J]. Electronic Measurement Technology, 2016, 39(6): 57-60.

[16] 李娟,吴谨,陈振学,等. 基于自学习的稀疏正则化图像超分辨率方法[J]. 仪器仪表学报,2015,36(1): 194-200.
LI J, WU J, CHEN ZH X, et al. Self-learning image

super-resolution method based on sparse represetation[J]. Chinese Journal of Scientific Instrument, 2015, 36(1): 194-200.
[17] DONG C, LOY C C, HE K, et al. Image super-resolution using deep convolutional networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2016, 38(2): 295-307.
[18] SONG X, DAI Y, QIN X. Deep depth super-resolution: learning depth super-resolution using deep convolutional neural network[C]. Asian Conference on Computer Vision, 2016: 360-376.
[19] NAIR V, HINTON G E. Rectified linear units improve restricted Boltzmann machines[C]. International Conference on Machine Learning, DBLP, 2010: 807-814.
[20] 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.
[21] 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]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 1874-1883.
[22] YANG J, WRIGHT J, HUANG T S, et al. Image super-resolution via sparse representation[J]. IEEE Transactions on Image Processing, 2010, 19(11): 2861-2873.
[23] 曲景影,孙显,高鑫. 基于CNN模型的高分辨率遥感图像目标识别[J]. 国外电子测量技术,2016,35(8): 45-50.
QU J Y, SUN X, GAO X, Remote sensing image target recognition based on CNN[J]. Foreign Electronic Measurement Technology, 2016, 35(8): 45-50.
[24] WANG Y, WANG L, WANG H, et al. End-to-end image super-resolution via deep and shallow convolutional networks[J]. Computer Vision and Pattern Recognition, 2016, arXiv:1607.07680.
[25] MAO X J, SHEN C, YANG Y B. Image restoration using convolutional auto-encoders with symmetric skip connections[J]. Computer Vision and Pattern Recognition, 2016, arXiv:1606.08921.
[26] RIEGLER G, RUTHER M, BISCHOF H. Atgv-net: Accurate depth super-resolution[C]. European Conference on Computer Vision. Springer International Publishing, 2016: 268-284.
[27] 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.
[28] CUN Y, BOTTOU L, BENGIO Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.
[29] SCHARSTEIN D, SZELISKI R, ZABIH R. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms[C]. IEEE Workshop on Stereo and Multi-Baseline Vision, 2001: 131-140.
[30] SCHARSTEIN D, SZELISKI R. High-accuracy stereo depth maps using structured light[C]. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, IEEE, 2003.
[31] SCHARSTEIN D, PAL C. Learning conditional random fields for stereo[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2007: 1-8.
[32] HIRSCHMULLER H, SCHARSTEIN D. Evaluation of cost functions for stereo matching[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2007: 1-8.
[33] RICHARDT C, STOLL C, DODGSON N A, et al. Coherent spatiotemporal filtering, upsampling and rendering of RGBZ videos[C]. Computer Graphics Forum, 2012: 247-256.
[34] 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.
[35] JIA Y, SHELHAMER E, DONAHUE J, et al. Caffe: Convolutional architecture for fast feature embedding[C]. Proceedings of the 22nd ACM International Conference on Multimedia, 2014: 675-678.
[36] LONG J, SHELHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015: 3431-3440.
[37] HE K, SUN J. Convolutional neural networks at constrained time cost[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015: 5353-5360.

作者简介



李伟,1993年出生,2015年于安徽建筑大学获得学士学位,现为合肥工业大学硕士研究生,主要研究方向为机器学习和图像处理。
E-mail:18256086350@163.com

Li Wei was born in 1993, received B. Sc. from Anhui Jianzhu University in 2015. And he is a M. Sc. candidate in Hefei University of Technology now. His main research direction is machine learning and image processing.

张旭东,1966年出生,2005年于中国科技大学获得博士学位,现任合肥工业大学教授,主要研究方向包括图像处理、机器学习和智能信息处理。
E-mail:xudong@hfut.edu.cn

Zhang Xudong was born in 1966, received Ph. D. from University of Science and Technology of China in 2005. And he is a professor in Hefei University of Technology. His research interests include image processing, machine learning and intelligent information processing.