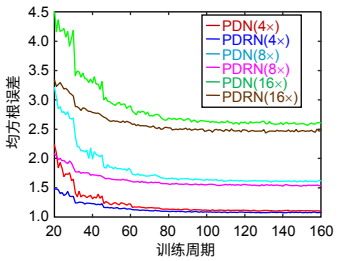
Opto-Electronic Engineering
光 电 工 程

Article

2019 年 , 第 46 卷 , 第 11 期

DOI: 10.12086/oee.2019.180587

级联金字塔结构的深度图超分辨率重建



付绪文，张旭东*，张　骏，孙　锐

合肥工业大学计算机与信息学院，安徽 合肥 230601

摘要：由于成像设备的限制，深度图往往分辨率较低。对低分辨率深度图进行上采样时，通常会造成深度图的边缘模糊。当上采样因子较大时，这种问题尤为明显。本文提出金字塔密集残差网络，实现深度图超分辨率重建。整个网络以残差网络为主框架，采用级联的金字塔结构对深度图分阶段上采样。在每一阶段，采用简化的密集连接块获取图像的高频残差信息，尤其是底层的边缘信息，同时残差结构中的跳跃连接分支获取图像的低频信息。网络直接以原始低分辨率深度图作为输入，以亚像素卷积层进行上采样操作，减少了运算复杂度。实验结果表明，该方法有效地解决了图像深度边缘的模糊问题，在定性和定量评价上优于现有方法。

关键词：深度图；超分辨率；金字塔；密集残差

中图分类号：TB872　　**文献标志码：**A

引用格式：付绪文，张旭东，张骏，等. 级联金字塔结构的深度图超分辨率重建[J]. 光电工程，2019，**46**(11): 180587

Depth map super-resolution with cascaded pyramid structure

Fu Xuwen, Zhang Xudong*, Zhang Jun, Sun Rui

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

Abstract: Due to the limitation of equipment, the resolution of depth map is low. Depth edges often become blurred when the low-resolution depth image is upsampled. In this paper, we present the pyramid dense residual network (PDRN) to efficiently reconstruct the high-resolution images. The network takes residual network as the main frame and adopts the cascaded pyramid structure for phased upsampling. At each pyramid level, the modified dense block is used to acquire high frequency residual, especially the edge features and the skip connection branch in the residual structure is used to deal with the low frequency information. The network directly uses the low-resolution depth image as the initial input of the network and the subpixel convolution layers is used for upsampling. It reduces the computational complexity. The experiments indicate that the proposed method effectively solves the problem of blurred edge and obtains great results both in qualitative and quantitative.

Keywords: depth map; super-resolution; pyramid; dense residual

Citation: Fu X W, Zhang X D, Zhang J, *et al.* Depth map super-resolution with cascaded pyramid structure[J]. *Opto-Electronic Engineering*, 2019, **46**(11): 180587

收稿日期：2018-11-14； 收到修改稿日期：2019-03-12

基金项目：国家自然科学基金资助项目(61471154，61876057)

作者简介：付绪文(1993-)，男，硕士研究生，主要从事智能信息处理、深度图像的研究。E-mail: fuxuwen@mail.hfut.edu.cn

通信作者：张旭东(1966-)，男，博士，教授，主要从事智能信息处理、机器视觉的研究。E-mail: xudong@hfut.edu.cn

1 引 言

近年来，随着三维成像技术的发展，深度信息得到了广泛的应用，例如虚拟现实、三维重建、人机交互[1]等。然而，由于硬件条件的限制，深度相机获取的图像分辨率较低，难以满足现实需要。比如常见的 PMD Camcube 相机，其分辨率仅有 204×204，即便是分辨率较高的 Microsoft Kinect 也只有 640×480。因此，如何从一幅低分辨率的深度图恢复出高分辨率的深度图，已经成为一个研究热点。

目前，计算机视觉领域的研究学者们提出了多种超分辨率重建算法[2-6]。早期的算法主要采用插值或滤波[4,7]，能够快速地获得重建结果。但是，由于可用线索的限制，其重建结果往往是不理想的。尤其是当上采样因子较大时，图像的纹理、边缘等细微结构丢失。这是因为图像的细微结构在有限的空间分辨率条件下，难以被完全表达[2]。简单的插值和滤波算法容易造成深度图的模糊，这种现象在深度不连续的区域表现得尤为明显。

为了解决图像模糊问题，文献[3-4,7-9]利用同场景的高分辨率彩色图来引导深度图上采样。此类方法都基于一个假设，即同一场景下的彩色图和深度图具有相同的结构。高分辨率彩色图的边缘被用来定位深度图的边缘。然而，两者的结构往往不是完全一致的，彩色图通常比深度图包含更多的纹理信息[8]。因此，该类方法通常会导致纹理过度复制。

当前，深度学习被广泛地应用在图像分类[10]、目标识别[11]、图像分割[12]等领域。考虑到卷积神经网络(convolutional neural network，CNN)强大的特征学习和映射能力，研究学者尝试用 CNN 完成超分辨率(super resolution，SR)任务。Dong 等[13]提出超分辨率卷积神经网络(super resolution convolution neural network，SRCNN)，第一次将 CNN 用于 SR 任务。此后，文献[14]基于稀疏编码理论，对 SRCNN 进行了改进。文献[15]则利用残差网络结构，搭建更深的网络模型用于超分辨率重建任务。这些方法同样存在问题：1) 部分网络的输入图像需要双三次插值的预处理过程：该预处理过程不仅增加了运算量，而且必然造成深度边缘的模糊；2) 网络直接以最后一层的特征进行超分辨率重建：深度图不同于彩色图，深度图代表距离信息，包含的纹理信息较少，主要表征了场景内物体的结构边缘。因此深度图的边缘是超分辨率重建的关键。直接以最后一层的特征信息进行重建，则忽视了深度图原始的边缘结构对重建结果的影响。

为了解决上述问题，本文提出一种新颖的端到端网络模型—金字塔密集残差网络(pyramid dense residual network，PDRN)用于深度图 SR。该网络以低分辨率深度图作为输入，避免插值预处理造成的边缘模糊。网络通过级联的金字塔模式对图片进行分阶段的上采样，利用各阶段获取的先验信息。每一个阶段具有相同的网络结构，且各阶段的上采样因子均为 2。每个阶段，采用简化的密集连接块[16]提取特征图，利用亚像素卷积层[17]对特征图进行上采样，并通过对残差结构的两个分支求和重建高分辨率深度图。采用更为鲁棒的 Charbonnier 损失函数[18]对网络各阶段的上采样结果进行约束。图 1 展示了上采样因子为 4 时整体的网络结构。

贡献：1) 本文采用两种跳跃连接：残差结构和简化的密集连接结构。残差结构将深度图的高低频信息分离；密集连接则加强了特征传递，尤其是对底层边缘特征的传递，使得图像重建时能够更加充分地利用边缘特征。2) 本文结合金字塔结构，对上采样过程进行多阶段约束。当上采样因子较大时，依然能够获得

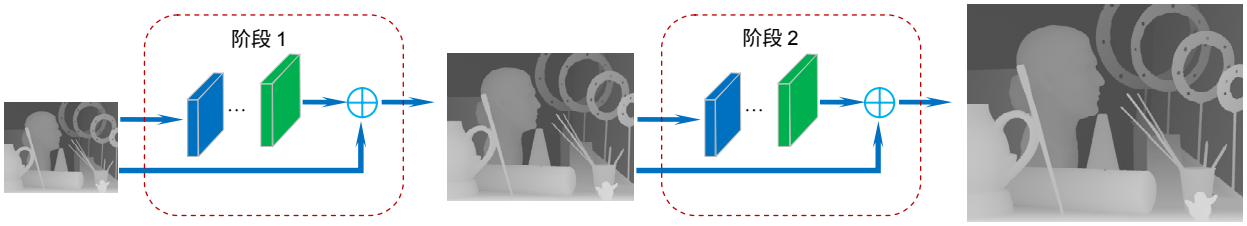


图 1 整体网络结构。该图是上采样因子为 4 时的结构图，包含 2 个阶段。每个阶段上采样因子为 2，一个阶段的输出作为下一阶段的输入

Fig. 1 The network structure. It is for upscale factor 4 that contains two stages. The upscale factor for each stage is 2 and the output of one stage is used as the input of the next stage

理想的重建结果。3) 由于每阶段网络结构完全一致，训练好的上采样因子为 2 的网络参数用来初始化上采样因子更大时的网络模型，从而极大地加速网络收敛。

2 相关工作

目前改善深度图分辨率的方法很多，简单归结为三类：1) 基于局部的方法；2) 基于全局的方法；3) 基于 CNN 的方法。

基于局部的方法：此类方法主要是基于滤波算法。Kopf 等[19]提出联合双边滤波器进行图像上采样，利用了图像的颜色信息，在一定程度上保留了图像边缘。Yang 等[20]依据低分辨率深度图求取代价体，并结合同场景彩色图像，在迭代过程中不断进行联合双边滤波，以得到更好的重建结果。文献[21]采用噪声敏感的联合双边滤波器处理上采样问题，以得到更小的深度误差。Lu 等[22]则利用同场景彩色图和深度图分割边缘的相关性来改善深度图的分辨率。文献[23]提出一种新颖的 SR 算法，并命名为改进的联合三边上采样，利用更多的彩色图线索使得结果具有清晰的边缘。然而此类方法利用的图像信息有限，深度图的边缘都存在不同程度的模糊问题。

基于全局的方法：此类方法通常是构建代价函数或者能量函数，从而将 SR 问题转化为最优化求解问题。Diebel 和 Thrun[24]首先使用马尔可夫随机场(Markov random field，MRF)模型，通过纹理的梯度对平滑项进行加权，建立彩色图和深度图之间的联系，有效地完成了深度图超分辨率重建任务。文献[25]基于最小二乘法框架，结合不同的加权项(包括分割、图像梯度、边缘显著性)和非局部均值算法(nonlocal means filtering，NLM)进行上采样。Aodha 等[26]则基于 MRF 模型，结合一种新颖的去噪算法来提高深度图的重建质量。基于广义总变分(total generalized variation，TGV)正则化，Ferstl 等[3]把同场景彩色图像作为深度图的边缘线索进行超分辨率，有效地保留了边缘信息。文献[27]建立彩色图像引导的自回归模型用于深度图恢复。Lei 等[28]考虑到视点合成质量，同时兼顾了深度的平滑性和纹理的相似性，从而得到更好的上采样结果。此类方法虽然保留了一定的边缘结构，但是运算复杂度较高，并且大多采用同场景的彩色图进行约束，容易发生纹理复制现象。

基于 CNN 的方法：当前，深度学习在计算机视觉方面取得了巨大的成功，例如图像分割、目标检测以及超分辨率等任务。文献[13]第一次提出端到端的超分辨率卷积神经网络(SRCNN)用于单幅图像的超分辨率，但是其网络结构过于简单，学习到的特征较为单一，而且网络收敛速度慢。Shi 等[17]提出亚像素卷积网络(efficient sub-pixel convolutional neural network，ESPCN)避免深度图上采样的预处理过程，减少了运算量的同时改善了重建结果。Hui 等[2]将彩色图片引导深度图上采样思想引入 CNN，提出多尺度引导卷积网络(multi-scale guided convolutional network，MSG-Net)进行深度图的超分辨率重建。文献[15]借助残差结构易训练的优势，提出一种非常深的卷积网络用于超分辨率(very deep convolutional for super resolution，VDSR)。基于金字塔模型，Lai 等[18]提出拉普拉斯金字塔网络(Laplacian pyramid super-resolution network，LapSRN)，采用级联的方式达到超分辨率的目的。但是上述方法中，部分网络采用插值算法进行图片预处理，在网络输入端已经存在边缘模糊现象，这在一定程度上提升了图像重建的难度。此外，上述网络全部以最后的网络特征图重建高分辨率图像，忽略了底层边缘特征对深度图重建的重要影响。

与上述网络模型相比，本文的网络主要有三点不同：1) 本文结合密集连接和残差结构，不仅加强了特征传递，而且加速了网络收敛。残差结构在学习过程中巧妙地将图像的高低频信息分离，重点对图像的高频细节信息进行学习，从而提高重建质量。密集连接则加强了底层边缘特征对重建结果的影响；2) 降低密集连接的密度，在跳跃连接结构之间增加卷积层以加深网络，同时每一次跳跃连接后增加特征图数量，使特征信息得到有效地传递和表达；3) 采用金字塔结构，以网络各阶段重建的深度图作为下一阶段的输入，通过对各个阶段上采样结果的多重约束，提高深度图重建的效果。同时由于各阶段网络结构完全一致，大大缩短了网络训练时间。

3 金字塔密集残差网络

本文采用的网络整体结构如图 1 所示，称为金字塔结构。该结构利用图像的多尺度特征，由粗到细的方式对图像进行多阶段上采样[29]。每个阶段具体的网络结构如图 2 所示。针对深度图纹理少、主要表征场景内物体结构边缘的特点，网络以残差网络[30]为主框架。残差结构能够分离图像的高低频信息，主分支、跳跃连接分支分别预测深度图的高、低频分量，其中高频分量对应深度图场景中物体的结构边缘。两个分支均采用亚像素卷积层[16]进行上采样。最后，残差结

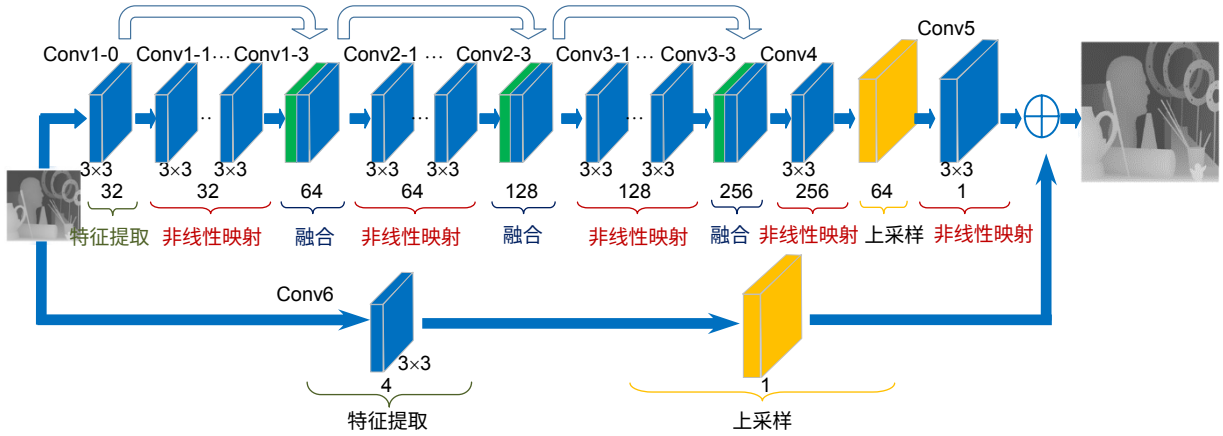


图 2 金字塔密集残差网络(PDRN)单阶段网络结构，上采样因子为 2。该结构包含四种运算：特征提取、非线性映射、融合、上采样

Fig. 2 The single stage of the pyramid dense residual network (PDRN) for upscale factor 2. It contains four operations: feature extraction, non-linear, fusion and upsampling

构的两个支路求和，即得到重建结果。同时残差网络能够防止梯度消失和爆炸，加速网络收敛。

图 3 显示了残差结构中两个分支和最终的重建结果图，以及对应的频谱分析图。子带残差结果对应的频谱图包含大量高频信息，而跳跃连接分支则主要包含图片的低频分量。该图表明本文的网络结构能够有效地分离图像的高低频信息，并重点学习图像的高频分量，以获得更精确的深度图边缘结构信息。

原始的密集连接块如图 4(a)所示，不同于传统的残差网络，本文采用一个简化的密集连接块[16]如图 4(b)所示，替换主分支的多层卷积层，把不同卷积层得到的特征图拼接在一起。在原始的密集连接块中，每一个卷积层之间都存在跳跃连接。本文采用的密集连接结构则减少了跳跃连接的密度，每 3 个卷积层存在一次跳跃连接，并且每进行一次跳跃连接，特征图的数量变为之前的 2 倍。采用密集连接结构，保证特征传递，减少信息丢失，并充分利用底层的边缘特征。密集连接结构的数学表示为：$x=[x_1,x_4,x_7,x_{10}]$，其中

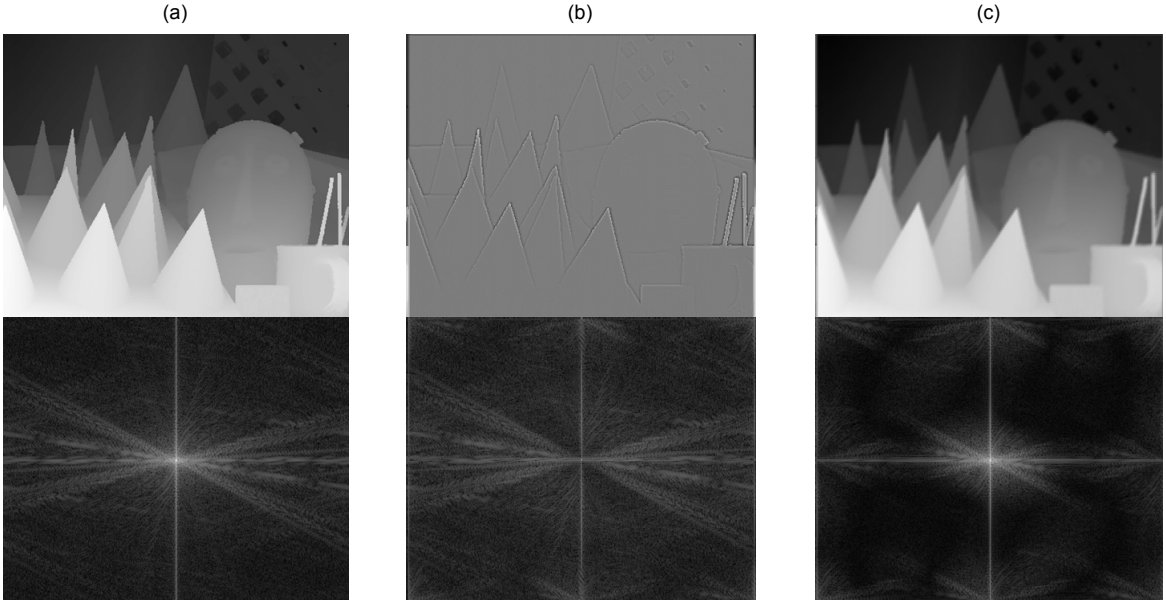


图 3 重建结果图及频谱分析图。(a) 最终结果图及对应的谱分析图；(b) 子带残差结果图及其频谱图；(c) 跳跃连接分支结果图及其频谱图

Fig. 3 Reconstruction results and spectrum analysis. (a) Final result and corresponding spectrum analysis; (b) Residual result and its spectrum analysis; (c) Skip connection result and its spectrum analysis

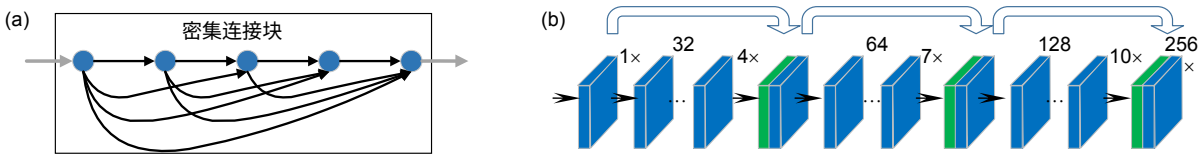


图 4 (a) 常见的密集连接块; (b) 本文采用的密集连接块，通过跳跃连接，特征图的数量由 32 增加到 256

Fig. 4 (a) The original dense block; (b) Our dense block that the number of feature maps is changed from 32 to 256 through three skip connections

$[x_1, x_4, x_7, x_{10}]$表示把卷积层 1、4、7、10 所产生的特征图拼接在一起。

整个网络以低分辨率深度图作为输入，以 s 个阶段($s=\log_2 U_f$)的残差网络构成金字塔结构实现图像重建，其中 U_f 表示上采样因子。定义 X 为低分辨率(LR)深度图。超分辨率任务的目标是由 X 预测高分辨率深度图 $F(X)$，使得 $F(X)$ 与真实的高分辨率深度图 Y 尽可能相似。网络需要学习 X 与 Y 之间的映射关系 F，其包含四种运算：

特征提取：该运算从原始的低分辨率图片 X 中提取高维特征向量。通常情况下，该运算得到的是图片底层边缘特征，类似于传统图像处理中的边缘检测。这些向量表征为一系列特征图的集合。

非线性映射：该运算将一个特征空间的特征图映射到更高层的特征空间。随着网络结构的加深，获取的特征图越抽象。

融合：类似于文献[16]所提到的，本文在某些卷积层与其后续卷积层之间建立跳跃连接。该运算将不同卷积层的特征向量拼接为一个更大的特征向量。这在一定程度上加强了特征传递，减少了特征传递过程中的信息丢失。

上采样：该运算采用亚像素卷积层[17]，将若干分辨率相同的特征图合为一幅更大的特征图，从而达到上采样的目的。

基于上述的四种运算，图 2 所示网络模型的数学描述如下：

特征提取：

$$F_1^0=\sigma(W_1^{(0)}*X+\boldsymbol{B}_1^{(0)}) \quad , \tag{1}$$

非线性映射：

$$F_j^k=\sigma(W_j^{(k)}*F_j^{(k-1)}+\boldsymbol{B}_j^{(k)}) \quad , \tag{2}$$

融合：

$$F_2^{(0)}=[F_1^{(3)},F_1^{(0)}] \quad , \tag{3}$$

融合：

$$F_3^{(0)}=[F_2^{(3)},F_2^{(0)}] \quad , \tag{4}$$

融合，非线性映射：

$$F_4=\sigma(W_4*[F_3^{(3)},F_3^{(0)}]+\boldsymbol{B}_4) \quad , \tag{5}$$

上采样，非线性映射：

$$F_5=\sigma(W_5*F_4^{\uparrow}+\boldsymbol{B}_5) \quad , \tag{6}$$

特征提取：

$$F_6=\sigma(W_6*X+\boldsymbol{B}_6) \quad , \tag{7}$$

上采样，重建：

$$F=F_5+F_6^{\uparrow} \quad 。 \tag{8}$$

上述各式中 $j=\{1,2,3\}, k=\{1,2,3\}$。运算符 $*$ 和 $\uparrow$ 分别表示卷积和亚像素卷积上采样。X 表示输入的低分辨率深度图。W 表示尺寸为 $n_{i-1}\times f_i\times f_i\times n_i$ 的卷积核权重(其中 n_{i-1} 和 n_i 分别表示第 $i-1$ 层和第 i 层的特征图数量，f_i 表示卷积核的大小)，$\boldsymbol{B}$ 表示一个 n_i 维的偏置向量。例如，$W_2^{(1)}$ 表示图 2 所示 Conv2-1 的卷积核。卷积层均采用参数校正线性单元(parametric rectified linear unit，PReLu)[31]作为激活函数用于非线性映射。

不同于大多数超分辨率任务中采用的损失函数：真实 HR 深度图与预测深度图之间的均方根误差(root mean square error, RMSE)。本文采用更鲁棒的 Charbonnier 损失函数[18]对各阶段的重建结果进行约束，提高图像重建的稳定性。通过对原始 HR 深度图下采样得到各个阶段真实的 HR 深度图 y_s 作为标签，对应的高分辨率深度图的预测值定义为 $\tilde{y}_s$。Charbonnier 损失函数如式(9)所示：

$$L(\tilde{y},y;\theta)=\frac{1}{N}\sum_{n=1}^{N}\sum_{s=1}^{S}l(\tilde{y}_s^{(n)}-y_s^n) \quad , \tag{9}$$

其中：$l(x)=\sqrt{x^2+\varepsilon^2}$，$N$ 表示每个批次的训练样本数，S 表示在金字塔结构中包含的阶段数。本文中将 ε 设

置为 1E-3。

上述损失函数是 RMSE 的改进版。本文的网络模型中，每个阶段都存在 RMSE 损失函数。整体的损失函数是由金字塔结构中所有阶段的 RMSE 损失函数求和得到。通过对金字塔结构中各阶段的结果进行约束，网络可以得到更稳定、更精确的重建结果。

4 实验结果与分析

将本文提出的方法在常用数据集上进行定性和定量的分析。首先，介绍网络训练过程中采用的训练集和测试集；之后介绍具体的训练细节；然后分析残差结构对超分辨率重建结果的影响；最后将本文的方法与现有的先进方法进行对比。

4.1 数据集

选取 Middlebury 数据集[32-34]中的 29 张深度图、RGBZ 数据集[7]的 10 张深度图和 ICL-NUIM[35]数据集的 45 张深度图组成包含 84 张图片的原始训练集。之后对深度图截取图像块，最终形成 80840 个图像块作为最终的训练集。选取 Middlebury 数据集和 Laser Scan 数据集作为测试集。采用两种方式进行数据增强：

1) 旋转，对原始数据集分别进行 90°、180°和 270°的旋转；

2) 翻转，在旋转的基础上，对图像进行水平翻转。所有的深度数据都被归一化为 $[-1,1]$。

4.2 训练细节

从原始图像中随机截取子图像块进行训练。相比较于直接使用大尺寸的图像，此种训练模式可以大大减少训练时间和内存消耗[12]，同时不会导致训练结果变差。与常见的超分辨率神经网络[13,15,17]相同，本文利用下采样获得低分辨率深度图像块。在每个训练批次，当上采样因子分别为 2 和 4、8、16 时，分别在同一幅图上截取 4 个 64×64、32×32、16×16 的图像块。实际训练过程中，为了进一步数据增强和防止过拟合，在每个图像块中添加了与图像方差正相关的随机噪声。

基于 tensorflow 深度学习框架搭建网络模型，采用 GTX 1080Ti 显卡进行网络训练。对于卷积层权重的初始化，统一采用文献[36]提出的方法。偏置初始化为 0。初始学习率设置为 5E-3，每经过 15 个训练周期，学习率变为之前的 $1/2$，共训练 180 个周期。特别地，在上述训练的基础上，采用不添加随机噪声的图像块进行网络调优。学习率重置为 5E-4，每间隔 15 个训练周期，学习率变为之前的 $1/10$，共训练 60 个周期。训练所采用的优化算法是自适应矩估计算法(adaptive moment estimation, Adam)[37]，其中两个动量参数分别设置为

$$\beta_1=0.9\ ,\quad \beta_2=0.9999\ 。$$

网络结构的参数设置：所有的卷积核尺寸 $f=3$，卷积核数量 $n_j^{(k)}=32\times 2^{(j-1)}$，$j=\{1,2,3\}$，$k=\{1,2,3\}$。例如 $n_2^{(3)}=32\times 2^{(2-1)}=64$ 为图 2 中 Conv2-3 的卷积核数量。

为了有效地评价重建结果，采用均方根误差(root mean square error，RMSE)和结构相似性(structural similarity index，SSIM)作为客观评价指标。RMSE 值越小，表示重建结果越好；SSIM 是指两幅图的结构相似程度，其值越接近 1，表明重建质量越好。

$$E_{\mathrm{RMSE}}=\sqrt{\frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}\left[X(i,j)-Y(i,j)\right]}\ ,\qquad (10)$$

$$M_{\mathrm{SSIM}}=\frac{(2\mu_x\mu_y+C_1)(2\sigma_{xy}+C_2)}{(\mu_x^2+\mu_y^2+C_1)(\sigma_x^2+\sigma_y^2+C_2)}\ ,\qquad (11)$$

其中：重建后的图像大小为 $M\times N$，X 和 Y 分别表示参考图像和重建后的图像；μ_x 和 σ_x 表示参考图像的灰度平均值和标准差；μ_y 和 σ_y 代表重建图像的灰度平均值和标准差；σ_{xy} 表示参考图像和重建图像间的协方差；C_1 和 C_2 为常数，避免分母为零。

4.3 残差结构的影响

为了分析残差结构对超分辨率重建结果的影响，在图 2 的基础上去掉跳跃连接分支，构成无残差结构的网络模型，并称其为金字塔密集网络(pyramid dense network，PDN)。表 1 显示了关于重建结果的定量评价数据。表中加粗的数据为最优结果，带有下划线的数据为次优结果。通过数据对比，网络中加入残差结构后，针对上采样因子为 2、4、8 时，RMSE 平均分别降低 0.39、0.14、0.15，SSIM 平均分别提升 8E-4、5E-4、7E-4。该实验证明，采用残差结构的模型获得的重建结果具有更低的 RMSE 和更高的 SSIM。

图 5 给出部分重建结果图，以此更直观地比较重建结果。从图中可以看出，最近邻插值和双三次插值算法得到的重建图像，其边缘出现明显的模糊和伪影，丢失大量的纹理细节。对比图 5(d)和图 5(e)，残差结构提高了网络的重建性能。图 6 显示了 PDN 和 PDRN 的训练收敛曲线，相比于 PDN，带有残差结构的 PDRN 收敛速度更快，并且 RMSE 更低。

表 1　有无残差结构在数据集 Laser Scan 的定量评价(RMSE/SSIM)

Table 1　Quantitative comparison (in RMSE and SSIM) on dataset Laser Scan

	Scan1			Scan2			Scan3		
	2×	4×	8×	2×	4×	8×	2×	4×	8×
Nearest	5.401/0.9820	8.567/0.9573	13.02/0.9176	4.251/0.9841	6.701/0.9630	9.992/0.9301	4.540/0.9842	8.112/0.9560	11.15/0.9218
Bicubic	4.215/0.9864	6.496/0.9689	10.09/0.9377	3.477/0.9873	5.234/0.9734	7.825/0.9505	4.063/0.9865	6.415/0.9658	8.930/0.9385
PDN	2.480/0.9917	3.754/0.9873	5.542/0.9791	2.106/0.9921	3.139/0.9880	4.455/0.9809	1.861/0.9951	2.889/0.9920	4.196/0.9866
PDRN	**2.170/0.9923**	**3.612/0.9878**	**5.391/0.9797**	**1.800/0.9927**	**3.041/0.9884**	**4.355/0.9815**	**1.303/0.9962**	**2.695/0.9926**	**4.006/0.9876**

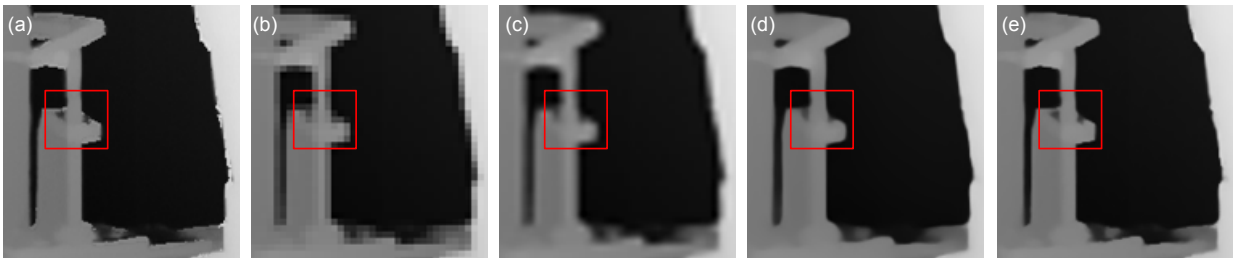


图 5　上采样为 4 的重建结果。(a) 真实深度图；(b) 最近邻插值；(c) 双三次插值；(d) PDN；(e) 本文方法

Fig. 5　Reconstruction results for scale factor 4. (a) Ground truth; (b) Nearest; (c) Bicubic; (d) PDN; (e) PDRN

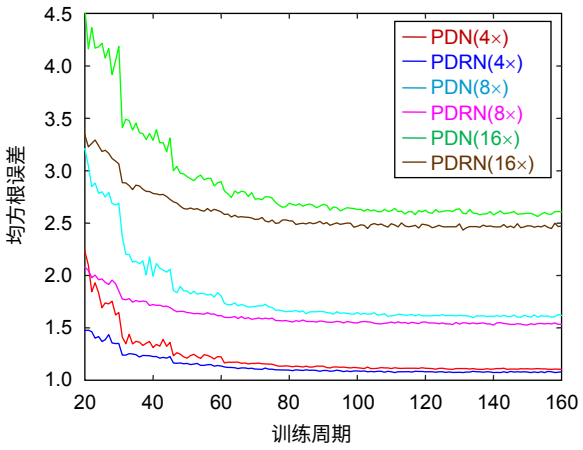


图 6　收敛曲线

Fig. 6　Convergent curve

4.4　分析对比

为了验证算法有效性，本文与 MRF[23]、Guided[4]、Edge[24]、TGV[3]、AP[26]、SRCNN[12]、VDSR[14]、MS-Net[2]、LapSRN[17]等方法进行定性和定量比较。所有方法的评价，基于 Middlebury 和 Laser Scan 数据集。本文将 Middlebury 数据集分为三部分：A(art，books，moebius)[6]，B(dolls，laundry，reindeer)[34]，C(cones，teddy，tsukuba)[35]和 D(Scan1，Scan2，Scan3)，采用 RMSE 和 SSIM 两个指标对上采样结果进行定量评价。表 2~表 7 中的数据，由相关文献的作者提供的图片计算得到(为保证实验公平性，SRCNN[13]、VDSR[14]、LapSRN[17]和本文的网络模型均采用相同数据集训练获得，且所有 HR 深度图转化为 8 位图)。每种评价的最优结果用加粗字体表示，次优结果用下划线标识。由于真实 HR 深度图分辨率低的限制，针对数据集 C 和 D 仅对上采样因子为 2、4 时的结果进行评价。

表 2 数据集 A 上重建结果的定量评价(RMSE)

Table 2 Quantitative comparison (in RMSE) on dataset A

	Art				Books				Moebius			
	2×	4×	8×	16×	2×	4×	8×	16×	2×	4×	8×	16×
Bicubic	2.5837	3.8565	5.5282	8.3759	1.0321	1.5794	2.2693	3.3652	0.9304	1.4006	2.0581	2.9702
MRF[24]	3.0192	3.6693	5.3349	8.3815	1.1976	1.5359	2.2026	3.4137	1.1665	1.4104	1.9905	2.9874
Guided[4]	2.8449	3.6686	4.8252	7.5978	1.1607	1.5646	2.0843	3.1856	1.0748	1.4029	1.8258	2.7592
Edge[25]	2.7502	3.4110	4.0320	6.0719	1.0611	1.5216	1.9754	2.7581	1.0501	1.3372	1.7863	2.3511
TGV[3]	2.9216	3.6474	4.6034	6.8204	1.2818	1.5841	1.9692	2.9503	1.1136	1.4302	1.8397	2.5216
AP[27]	1.7909	2.8513	3.6943	5.9113	1.3280	1.5297	1.8394	2.9187	0.8704	1.0311	1.5505	2.5728
SRCNN[13]	1.1338	2.0175	3.8293	7.2717	0.5231	0.9356	1.7268	3.1006	0.5374	0.9132	1.5790	2.6896
VDSR[15]	1.3242	2.0710	3.2433	6.6622	0.4883	0.8296	1.2878	2.1433	0.5441	0.8341	1.2952	2.1590
MS-Net[2]	0.8131	**1.6352**	2.7697	**5.8040**	0.4180	0.7404	1.0770	1.8165	0.4133	0.7448	1.1384	1.9151
LapSRN[18]	0.7644	2.3796	3.3389	6.1110	0.4443	0.9423	1.3430	1.9523	0.4300	0.9382	1.3388	2.0450
PDRN	**0.6233**	1.6634	**2.6943**	5.8083	**0.3875**	**0.7313**	**1.0668**	**1.7323**	**0.3723**	**0.7347**	**1.0767**	**1.8567**

表 3 数据集 A 上重建结果的定量评价(SSIM)

Table 3 Quantitative comparison (in SSIM) on dataset A

	Art				Books				Moebius			
	2×	4×	8×	16×	2×	4×	8×	16×	2×	4×	8×	16×
Bicubic	0.9868	0.9679	0.9433	0.9254	0.9956	0.9900	0.9835	0.9789	0.9950	0.9888	0.9811	0.9761
MRF[24]	0.9833	0.9749	0.9570	0.9371	0.9945	0.9916	0.9865	0.9811	0.9929	0.9901	0.9846	0.9792
Guided[4]	0.9830	0.9710	0.9579	0.9476	0.9945	0.9906	0.9865	0.9833	0.9934	0.9894	0.9853	0.9812
Edge[25]	0.9870	0.9797	0.9725	0.9605	0.9956	0.9918	0.9877	0.9846	0.9945	0.9910	0.9868	0.9840
TGV[3]	0.9858	0.9783	0.9668	0.9535	0.9945	0.9919	0.9886	0.9843	0.9941	0.9907	0.9860	0.9819
AP[27]	0.9952	0.9871	0.9798	0.9586	0.9961	0.9942	0.9909	0.9849	0.9972	0.9950	0.9906	0.9833
VDSR[15]	0.9964	0.9916	0.9801	0.9457	0.9985	0.9966	0.9932	0.9866	0.9980	0.9958	0.9913	0.9823
MS-Net[2]	0.9983	0.9941	0.9851	0.9557	0.9990	**0.9973**	0.9949	0.9894	0.9988	0.9965	0.9928	0.9851
LapSRN[18]	0.9984	0.9888	0.9791	0.9584	0.9986	0.9958	0.9931	0.9891	0.9986	0.9949	0.9912	0.9852
PDRN	**0.9988**	**0.9945**	**0.9871**	**0.9601**	**0.9990**	0.9971	**0.9949**	**0.9898**	**0.9988**	**0.9965**	**0.9935**	**0.9863**

表 4 数据集 B 上重建结果的定量评价(RMSE)

Table 4 Quantitative comparison (in RMSE) on dataset B

	Dolls				Laundry				Reindeer			
	2×	4×	8×	16×	2×	4×	8×	16×	2×	4×	8×	16×
Bicubic	0.9433	1.3356	1.8811	2.6451	1.6142	2.4077	3.4520	5.0923	1.9382	2.8086	3.9857	5.8591
Edge[25]	0.9713	1.3217	1.7750	2.4341	1.5525	2.1316	2.7698	4.1581	2.2674	2.4067	2.9873	4.2941
TGV[3]	1.1467	1.3869	1.8854	3.5878	1.9886	2.5115	3.7570	6.4066	2.4068	2.7115	3.7887	7.2711
AP[27]	1.1893	1.3888	1.6783	2.3456	1.7154	2.2553	2.8478	4.6564	1.8026	2.4309	2.9491	4.0877
SRCNN[13]	0.5814	0.9467	1.5185	2.4452	0.6353	1.1761	2.4306	4.5795	0.7658	1.4997	2.8643	5.2491
VDSR[15]	0.6308	0.8966	1.3148	2.0912	0.7249	1.1974	1.8392	3.2061	1.0051	1.5058	2.2814	4.1759
MS-Net[2]	0.4813	**0.7835**	1.2049	1.8627	0.4749	**0.8842**	1.6279	3.4353	0.5563	1.1068	1.9719	3.9215
LapSRN[18]	0.4942	1.0111	1.4194	1.9954	0.4617	1.3552	1.9314	2.5137	0.5075	1.7402	2.4538	3.0737
PDRN	**0.4418**	0.8317	**1.1916**	**1.8230**	**0.3897**	0.9240	**1.4207**	**2.2084**	**0.4149**	**1.1047**	**1.7392**	**2.6655**

表 5 数据集 B 上重建结果的定量评价(SSIM)

Table 5 Quantitative comparison (in SSIM) on dataset B

	Dolls				Laundry				Reindeer			
	2×	4×	8×	16×	2×	4×	8×	16×	2×	4×	8×	16×
Bicubic	0.9948	0.9893	0.9827	0.9783	0.9926	0.9833	0.9717	0.9635	0.9930	0.9846	0.9737	0.9642
Edge[25]	0.9950	0.9911	0.9876	0.9849	0.9938	0.9892	0.9850	0.9774	0.9897	0.9902	0.9863	0.9819
TGV[3]	0.9934	0.9902	0.9853	0.9770	0.9882	0.9820	0.9714	0.9551	0.9910	0.9874	0.9798	0.9690
AP[27]	0.9925	0.9901	0.9876	0.9842	0.9942	0.9905	0.9876	0.9757	0.9931	0.9891	0.9860	0.9794
VDSR[15]	0.9976	0.9953	0.9910	0.9457	0.9979	0.9835	0.9878	0.9762	0.9977	0.9953	0.9906	0.9773
MS-Net[2]	**0.9987**	**0.9964**	0.9921	0.9851	0.9989	0.9965	0.9899	0.9751	0.9989	0.9968	0.9929	0.9814
LapSRN[18]	0.9982	0.9941	0.9901	0.9851	0.9988	0.9942	0.9892	0.9858	0.9987	0.9938	0.9896	0.9868
PDRN	0.9985	0.9958	**0.9922**	**0.9860**	**0.9990**	**0.9970**	**0.9937**	**0.9872**	**0.9990**	**0.9968**	**0.9942**	**0.9887**

表 6 数据集 C 上重建结果的定量评价(RMSE/SSIM)

Table 6 Quantitative comparison (in RMSE/SSIM) on dataset C

	Cones		Teddy		Tsukuba	
	2×	4×	2×	4×	2×	4×
Bicubic	2.5422/0.9813	3.8666/0.9583	1.9610/0.9844	2.8583/0.9665	5.8201/0.9694	8.5638/0.9277
Edge[25]	2.8497/0.9699	6.5447/0.9420	2.1850/0.9767	4.3366/0.9553	6.8869/0.9320	12.123/0.8981
Ferstl[38]	2.1850/0.9866	3.4977/0.9645	1.6941/0.9884	2.5966/0.9716	5.3252/0.9766	7.5356/0.9413
Xie[39]	2.7338/0.9633	4.4087/0.9319	2.4911/0.9625	3.2768/0.9331	6.3534/0.9464	9.7765/0.8822
Song[40]	1.4356/0.9989	2.9789/0.9783	1.1974/0.9918	1.8006/0.9831	2.9841/0.9905	6.1422/0.9666
SRCNN[13]	1.4842/0.9965	3.5856/0.9672	1.1702/0.9923	1.9857/0.9820	3.2753/0.9879	7.9391/0.9587
VDSR[15]	1.7150/0.9917	2.9808/0.9797	1.2203/0.9925	1.8591/0.9836	3.7684/0.9896	5.9175/0.9686
MS-Net[2]	1.1005/0.9951	2.7659/0.9817	0.8204/0.9953	1.5283/0.9865	2.4536/0.9934	4.9927/0.9740
LapSRN[18]	1.0182/0.9958	3.1994/0.9755	0.8570/0.9951	2.0820/0.9802	2.0822/0.9960	6.2983/0.9649
PDRN	**0.8556/0.9963**	**2.6049/0.9837**	**0.7359/0.9959**	**1.6421/0.9860**	**1.8128/0.9974**	**4.9136/0.9798**

表 7 数据集 D 上重建结果的定量评价(RMSE/SSIM)

Table 7 Quantitative comparison (in RMSE/SSIM) on dataset D

	Scan1		Scan2		Scan3	
	2×	4×	2×	4×	2×	4×
Bicubic	4.2153/0.9864	6.4958/0.9689	3.4766/0.9873	5.2335/0.9734	4.0629/0.9865	6.4149/0.9658
Xie[39]	-	9.1935/0.9781	-	7.4148/0.9791	-	8.9093/0.9680
VDSR[15]	2.7391/0.9911	3.9732/0.9865	2.2883/0.9919	3.2704/0.9878	1.4128/0.9960	3.0617/0.9912
MS-Net[2]	2.7502/0.9902	3.7618/0.9836	2.1329/0.9917	3.4159/0.9863	1.4296/0.9955	3.1048/0.9914
LapSRN[18]	2.3995/0.9913	4.2889/0.9834	1.9860/0.9920	3.5537/0.9852	1.4702/0.9957	3.6258/0.9878
PDRN	**2.1698/0.9923**	**3.6122/0.9878**	**1.8003/0.9927**	**3.0407/0.9884**	**1.3034/0.9962**	**2.6953/0.9926**

表中的数据显示，本文方法(PDRN)性能优于当前先进的方法。对比传统与深度学习两类不同的算法，受益于深度学习在图像处理领域优越的表现，SRCNN 的重建结果优于传统的超分辨率重建算法。SRCNN、MS-Net 和 LapSRN 都是针对单幅图像的超分辨率重建网络模型，但是 MS-Net 和 LapSRN 的性能优于 SRCNN。原因在于 MS-Net 和 LapSRN 直接采用低分辨率深度图作为原始输入，避免了双三次插值预处理

过程带来的噪声。VDSR 则利用残差结构建立更深层的网络模型，从而得到性能的提升。本文提出的方法(PDRN)总体表现优于 MS-Net、VDSR 和 LapSRN。因为在本文网络结构结合残差结构和密集连接两种跳跃连接，避免了信息丢失，使特征得到充分利用。

图 7 和图 8 分别显示了在 Middlebury 和 Laser Scan 数据集上，不同方法的重建结果，同时对部分图像块放大显示。图中可以看出，TGV[3]、Xie[39]方法的效果最差，图像边缘发生严重模糊，并且出现纹理复制问题；AP[27]方法对于纹理丰富的区域重建结果比较理想，而当图像中缺乏纹理时，图像发生边缘模糊；MS-Net[2]和 LapSRN[18]方法重建结果较好，但是在狭窄的边缘依然存在模糊问题。本文的方法则有效地处理了边缘模糊问题，使重建的深度图具有更好的视觉效果。通过对比，本文提出的方法性能更好，得到的 HR 深度图具有更加清晰的边缘。

由于网络采用级联的金字塔结构，网络各阶段的结构完全一致。因此在网络训练过程中，首先训练上采样因子为 2 的网络模型，之后通过预训练的该网络模型，分别初始化上采样因子为 4、8、16 时各阶段的网络参数，最后只需要对网络模型进行调优即可。表 8 显示了采用预训练模型进行网络初始化的优势。数

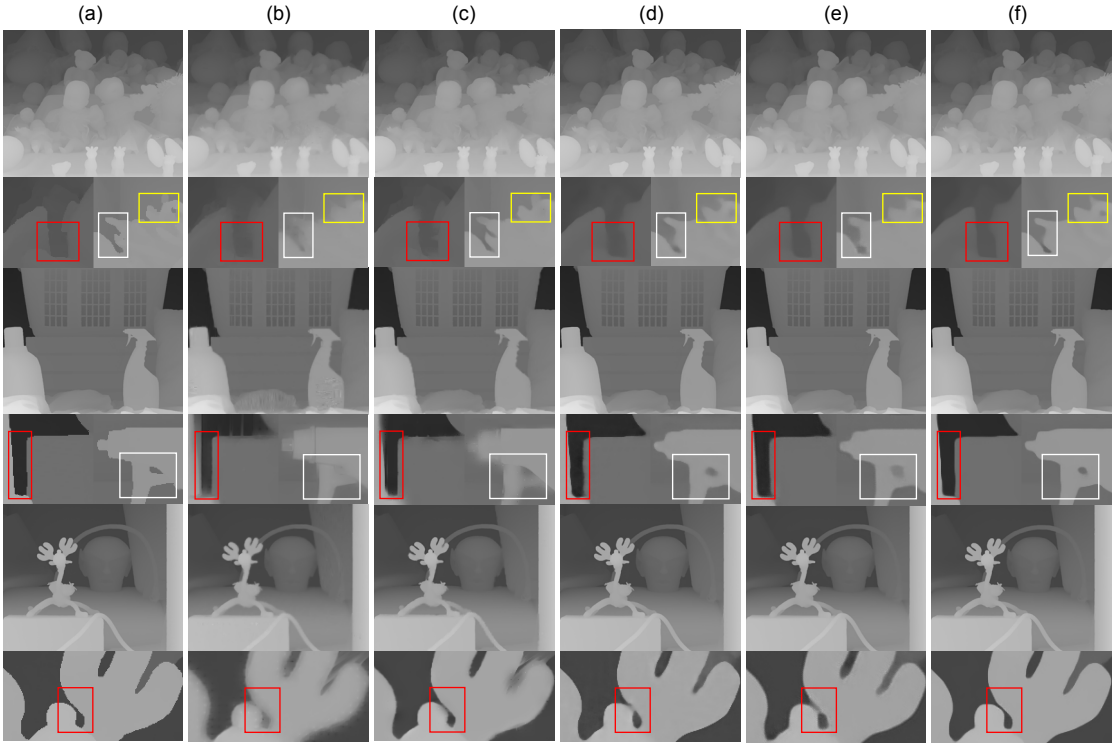


图 7 上采样深度图。(a) 真实深度图；(b) TGV[3]；(c) AP[27]；(d) MS-Net[2]；(e) LapSRN[18]；(f) 本文方法

Fig. 7 Upsampling depth map. (a) Ground truth; (b) TGV[3]; (c) AP[27]; (d) MS-Net[2]; (e) LapSRN[18]; (f) Ours

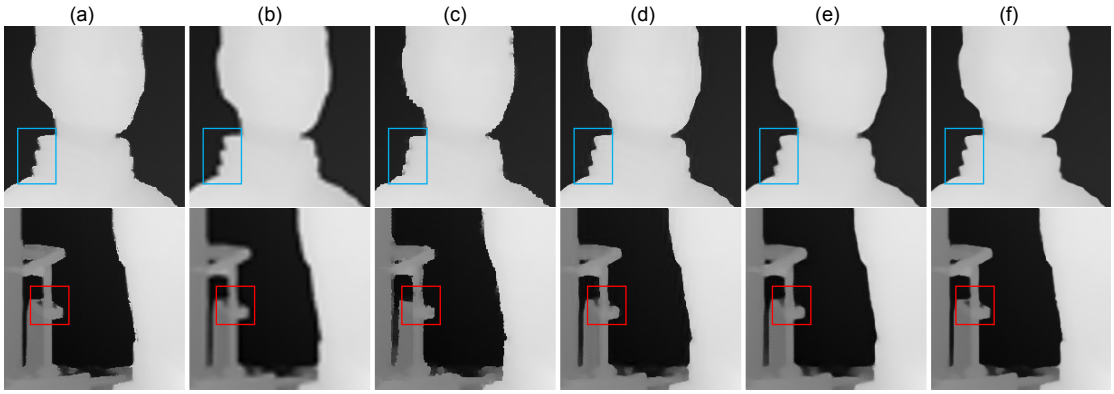


图 8 上采样深度图。(a) 真实深度图；(b) Bicubic；(c) Xie[39]；(d) MS-Net[2]；(e) LapSRN[18]；(f) 本文方法

Fig. 8 Upsampling depth map. (a) Ground truth; (b) Bicubic; (c) Xie[39]; (d) MS-Net[2]; (e) LapSRN[18]; (f) Ours

表 8 训练时间(小时)

Table 8 Training time(hour)

方法	4×	8×	16×
SRCNN[13]	12.35	12.65	13.16
VDSR[15]	4.65	4.59	4.86
LapSRN[18]	4.93	5.08	5.13
本文(标准初始化)	5.97	6.53	6.80
本文(预训练初始化)	1.49	1.63	1.70

据表明，由于 SRCNN 未采用残差结构，网络收敛最慢；VDSR 和 LapSRN 利用残差结构，加快网络收敛速度。本文在残差网络的基础上，采用预训练模型初始化网络参数，网络训练时间缩短为标准初始化方法的1/4。

5 结 论

为了解决深度图超分辨率过程中出现的边缘模糊和信息丢失问题，搭建了金字塔密集残差网络(PDRN)实现单幅深度图的超分辨率。通过金字塔结构和更为鲁棒的损失函数，充分利用图像的多尺度特征，对重建结果进行多层约束，得到更好的重建结果。残差结构用于分离深度图的高低频分量，对各分量进行针对性学习，同时也起到加速收敛的作用。考虑到深度图超分辨率任务中边缘重建的重要性，本文采用简化的密集连接块，充分利用卷积神经网络底层的边缘特征，并且密集连接块的特征图数量随着网络的加深而增加，加强了特征传递。实验的定性和定量评价表明：在视觉质量和重建精度两个方面，本文提出的方法较其他方法有了一定的提升。本文使用的数据为公用的合成数据集，未充分考虑实际深度图中的空洞以及光线导致的像素深度值缺失等问题。因此下一步工作将研究在上述情况下，如何有效地实现深度图的超分辨率重建。

参考文献

[1] Ruiz-Sarmiento J R, Galindo C, Gonzalez J. Improving human face detection through TOF cameras for ambient intelligence applications[C]//*Proceedings of the 2nd International Symposium on Ambient Intelligence*, 2011: 125–132.

[2] 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*, 2016: 353–369.

[3] Ferstl D, Reinbacher C, Ranftl R, *et al*. Image guided depth upsampling using anisotropic total generalized variation[C]//*Proceedings of 2013 IEEE International Conference on Computer Vision*, 2013: 993–1000.

[4] He K M, Sun J, Tang X O. Guided image filtering[C]//*Proceedings of the 11th European Conference on Computer Vision*, 2010: 1–14.

[5] Wang R G, Wang Q H, Yang J, *et al*. Image super-resolution reconstruction by fusing feature classification and independent dictionary training[J]. *Opto-Electronic Engineering*, 2018, **45**(1): 170542.
汪荣贵, 汪庆辉, 杨娟, 等. 融合特征分类和独立字典训练的超分辨率重建[J]. 光电工程, 2018, **45**(1): 170542.

[6] Wang F, Wang W, Qiu Z L. A single super-resolution method via deep cascade network[J]. *Opto-Electronic Engineering*, 2018, **45**(7): 170729.
王飞, 王伟, 邱智亮. 一种深度级联网络结构的单帧超分辨重建算法[J]. 光电工程, 2018, **45**(7): 170729.

[7] 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.

[8] Shen X Y, Zhou C, Xu L, *et al*. Mutual-structure for joint filtering[C]//*Proceedings of 2015 IEEE International Conference on Computer Vision*, 2015: 3406–3414.

[9] Kiechle M, Hawe S, Kleinsteuber M. A joint intensity and depth co-sparse analysis model for depth map super-resolution[C]//*Proceedings of 2013 IEEE International Conference on Computer Vision*, 2013: 1545–1552.

[10] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks[C]//*Proceedings of the 25th International Conference on Neural Information Processing Systems*, 2012: 1097–1105.

[11] Ren S Q, He K M, Girshick R, *et al*. Faster R-CNN: towards real-time object detection with region proposal networks[C]//*Proceedings of the 28th International Conference on Neural Information Processing Systems*, 2015: 91–99.

[12] 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.

[13] 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.

[14] Wang Z W, Liu D, Yang J C, *et al*. Deep networks for image super-resolution with sparse prior[C]//*Proceedings of IEEE International Conference on Computer Vision*, 2015: 370–378.

[15] 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*, 2016: 1646–1654.

[16] Huang G, Liu Z, Van Der Maaten L, *et al*. Densely connected

convolutional networks[C]//*Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition*, 2017: 2261–2269.

[17] Shi W Z, Caballero J, Huszár F, *et al*. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network[C]//*Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition*, 2016: 1874–1883.

[18] Lai W S, Huang J B, Ahuja N, *et al*. Deep laplacian pyramid networks for fast and accurate super-resolution[C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017: 5835–5843.

[19] Kopf J, Cohen M F, Lischinski D, *et al*. Joint bilateral upsampling[J]. *ACM Transactions on Graphics*, 2007, **26**(3): 96.

[20] Yang Q, Yang R, Davis J, *et al*. Spatial-depth super resolution for range images[C]//*Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition*, 2007: 1–8.

[21] Chan D, Buisman H, Theobalt C, *et al*. A noise-aware filter for real-time depth upsampling[C]//*Proceedings of Workshop on Multi-camera and Multi-modal Sensor Fusion Algorithms and Applications*, 2008.

[22] Lu J J, Forsyth D. Sparse depth super resolution[C]//*Proceedings of 2005 IEEE Conference on Computer Vision and Pattern Recognition*, 2015: 2245–2253.

[23] Yuan L, Jin X, Li Y G, *et al*. Depth map super-resolution via low-resolution depth guided joint trilateral up-sampling[J]. *Journal of Visual Communication and Image Representation*, 2017, **46**: 280–291.

[24] Diebel J, Thrun S. An application of Markov random fields to range sensing[C]//*Advances in Neural Information Processing Systems*, 2005: 291–298.

[25] 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*, 2011: 1623–1630.

[26] 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*, 2012: 71–84.

[27] Yang J Y, Ye X C, Li K, *et al*. Color-guided depth recovery from RGB-D data using an adaptive autoregressive model[J]. *IEEE Transactions on Image Processing*, 2014, **23**(8): 3443–3458.

[28] Lei J J, Li L L, Yue H J, *et al*. Depth map super-resolution considering view synthesis quality[J]. *IEEE Transactions on Image Processing*, 2017, **26**(4): 1732–1745.

[29] Denton E, Chintala S, Szlam A, *et al*. Deep generative image models using a Laplacian pyramid of adversarial networks[C]//*Proceedings of the 28th International Conference on Neural Information Processing Systems*, 2015: 1486–1494.

[30] He K M, Zhang X Y, Ren S Q, *et al*. Deep residual learning for image recognition[C]//*Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition*, 2016: 770–778.

[31] 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*, 2015: 1026–1034.

[32] Scharstein D, Szeliski R. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms[J]. *International Journal of Computer Vision*, 2002, **47**(1–3): 7–42.

[33] Scharstein D, Pal C. Learning conditional random fields for stereo[C]//*Proceedings of 2007 IEEE Conference on Computer Vision and Pattern Recognition*, 2007: 1–8.

[34] Scharstein D, Hirschmüller H, Kitajima Y, *et al*. High-resolution stereo datasets with subpixel-accurate ground truth[C]//*Proceedings of the 36th German Conference on Pattern Recognition*, 2014: 31–42.

[35] 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*, 2014: 1524–1531.

[36] Glorot X, Bengio Y. Understanding the difficulty of training deep feedforward neural networks[C]//*Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, 2010: 249–256.

[37] Kingma D, Ba J. Adam: a method for stochastic optimization[C]//*Proceedings of the 3rd International Conference on Learning Representations*, 2014.

[38] Ferstl D, Rüther M, Bischof H. Variational depth superresolution using example-based edge representations[C]//*Proceedings of 2015 IEEE International Conference on Computer Vision*, 2015: 513–521.

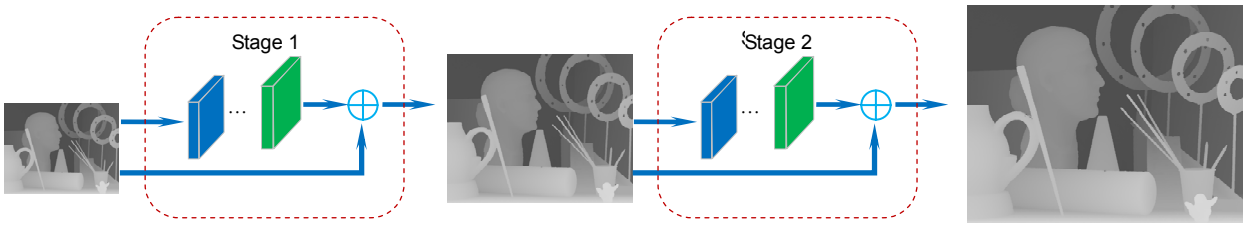
[39] Xie J, Feris R S, Sun M T. Edge guided single depth image super resolution[C]//*Proceedings of 2014 IEEE International Conference on Image Processing*, 2014: 3773–3777.

[40] Song X B, Dai Y C, Qin X Y. Deep depth super-resolution: learning depth super-resolution using deep convolutional neural network[C]//*Proceedings of the 13th Asian Conference on Computer Vision*, 2017: 360–376.

Depth map super-resolution with cascaded pyramid structure

Fu Xuwen, Zhang Xudong*, Zhang Jun, Sun Rui

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



The network structure

Overview: With the development of science and technology, depth information is gradually applied to various fields of society, such as face recognition, virtual reality and so on. However, due to the limitation of hardware conditions such as sensors, the resolution of the depth images is too low to meet the requirements of the reality. Depth map super- resolution has been an important research area in the field of computer vision. Early methods adopt interpolation, filtering. Although these methods are simple and fast in running, the clues used by these methods are limited and the results are not ideal. The details of the depth map is lost, especially when the upsampling factor is large. To address the above issue, intensity images are used to guide the depth map super-resolution. But when the local structures in the guidance and depth images are not consistent, these techniques may cause the over-texture transferring problem. At present, convolutional neural networks are widely used in computer vision because of its powerful feature representation ability. Several models based on convolutional neural networks have achieved great success in single image super-resolution. To solve the problems of edge blurring and over-texture transferring in depth super-resolution, we propose a new framework to achieve single depth image super-resolution based on the pyramid dense residual network (PDRN). The PDRN directly uses the low-resolution depth image as the input of the network and doesn't require the pre-processing. This can reduce the computational complexity and avoid the additional error. The network takes the residual network as the backbone model and adopts the cascaded pyramid structure for phased upsampling. The result of each pyramid stage is used as the input of the next stage. At each pyramid stage, the modified dense block is used to acquire high frequency residual, and subpixel convolution layer is used for upsampling. The dense block enhances transmission of feature and reduces information loss. Therefore, the network can reconstruct high resolution depth maps using different levels of features. The residual structure is used to shorten the time of convergence and improve the accuracy. In addition, the network adopts the Charbonnier loss function to train the network. By constraining the results of each stage, the training network can get more stable results. Experiments show that the proposed network can avoid edge blurring, detail loss and over-texture transferring in depth map super-resolution. Extensive evaluations on several datasets indicate that the proposed method obtains better performance comparing to other state-of-the-art methods.

Citation: Fu X W, Zhang X D, Zhang J, *et al*. Depth map super-resolution with cascaded pyramid structure[J]. *Opto-Electronic Engineering*, 2019, **46**(11): 180587

Supported by National Natural Science Foundation of China (61471154, 61876057)
* E-mail: xudong@hfut.edu.cn