·图形图像处理·　　文章编号：1000-3428(2017)11-0252-09　　文献标志码：A　　中图分类号：TP391.41

基于深度学习的深度图超分辨率采样

王晓晖，盛　斌，申瑞民

（上海交通大学 计算机科学与工程系，上海 201100）

摘　要：在深度图像采集场景下，为利用场景高分辨色彩图进行超分辨率上采样，提出一种采用卷积神经网络自适应学习局部滤波器核的算法，通过同时应用稠密/高分辨率颜色信息和稀疏/低分辨率深度信息全面提取场景信息。在 Middlebury 和 ToFMark 数据集上的实验结果表明，与传统深度超分辨率算法相比，提出的算法能够取得较好的超分辨率结果，尤其在颜色和深度的边缘、纹理不匹配区域，具有更好的鲁棒性。

关键词：深度超分辨率；上采样；滤波；深度学习；卷积神经网络；立体视觉

中文引用格式：王晓晖，盛　斌，申瑞民. 基于深度学习的深度图超分辨率采样[J]. 计算机工程，2017，43(11)：252-260.
英文引用格式：WANG Xiaohui，SHENG Bin，SHEN Ruimin. Deep Depth Graph Super Resolution Sampling Based on Depth Learning[J]. Computer Engineering，2017，43(11)：252-260.

Deep Depth Graph Super Resolution Sampling Based on Depth Learning

WANG Xiaohui，SHENG Bin，SHEN Ruimin

（Department of Computer Science and Engineering，Shanghai Jiaotong University，Shanghai 201100，China）

【Abstract】 In the scene of depth image acquisition，in order to use the high-resolution color map of the scene for superresolution upper sampling，this paper proposes an adaptive learning algorithm for local filter kernels using convolutional neural network. It utilizes both the dense/high-resolution color and the sparse/low-resolution depth information to extract the scene informationentirely. Experimental results on the Middlebury and ToFMark datasets show that，compared with traditional depth superresolution algorithms，the proposed algorithm is capable of obtaining the best super-resolution results. Especially in the color and depth edge as well as the texture mismatch region，it has better robustness.

【Key words】 Depth Super Resolution（DSR）；upsampling；filtering；depth learning；Convolutional Neural Network（CNN）；stereo vision

DOI：10.3969/j.issn.1000-3428.2017.11.041

0　概述

可靠的深度信息是很多计算机应用的重要信息来源，例如交互式自由视点视频[1]、三维重建[2]、语义场景分析[3]、增强现实、人机交互[4]和人体姿态识别[5]。最近，新的商业 3D 摄像机（包括基于传播延迟技术和基于结构光技术）改进了深度图像的采集，但是这些深度设备仍然受限于较低的分辨率。

在 SRCNN 算法启发下，本文提出一种将基于学习方法和基于保边滤波方法相结合的深度超分辨率（Depth Super Resolution，DSR）算法。它利用卷积神经网络从外部数据中学习滤波器核，对低分辨率深度图做滤波，从而得到超分辨率结果。更具体地说，卷积神经网络学习的是一个从低分辨率深度图像块、高分辨率颜色图像块到局部滤波器核的映射函数，卷积神经网络的应用实现同时从颜色信息和深度信息更全面地提取场景信息。

1　背景介绍

DSR 技术是一个很实用的解决办法，利用与深度图相匹配的高分辨颜色图获得辅助信息。这些算法采用了手工设计的具有保边性能的滤波器核，将局部颜色信息传播到深度图。此类方法建立在颜色图和深度图中边缘总是匹配的假设上，因此当这个假设被违背时，包括边缘模糊、纹理复制的上采样痕迹就会出现。如图 1 所示，图 1(c)中的保边滤波器（如联合双边滤波器）的滤波器核与图 1(d)中的最

基金项目：国家自然科学基金(61671290)。
作者简介：王晓晖(1992—)，男，硕士，主研方向为计算机视觉、机器学习；盛　斌(通信作者)，副教授；申瑞民，教授。
收稿日期：2016-11-16　　**修回日期：**2016-12-19　**E-mail：**wxhcr7@163.com

佳滤波器核有巨大差异。其中,最佳滤波器核是以图 1(f)中所示的场景真实高分辨深度图作为指导图用联合双边滤波器计算得到,在实际应用中无法获得。图中颜色值代表滤波器核权重,浅色代表高权重,深色代表低权重。如图中最后 2 行所示,当颜色边缘与深度边缘不一致时,联合双边滤波器对应的滤波器核与最佳滤波器核差异较大,而本文提出的滤波器核更接近。

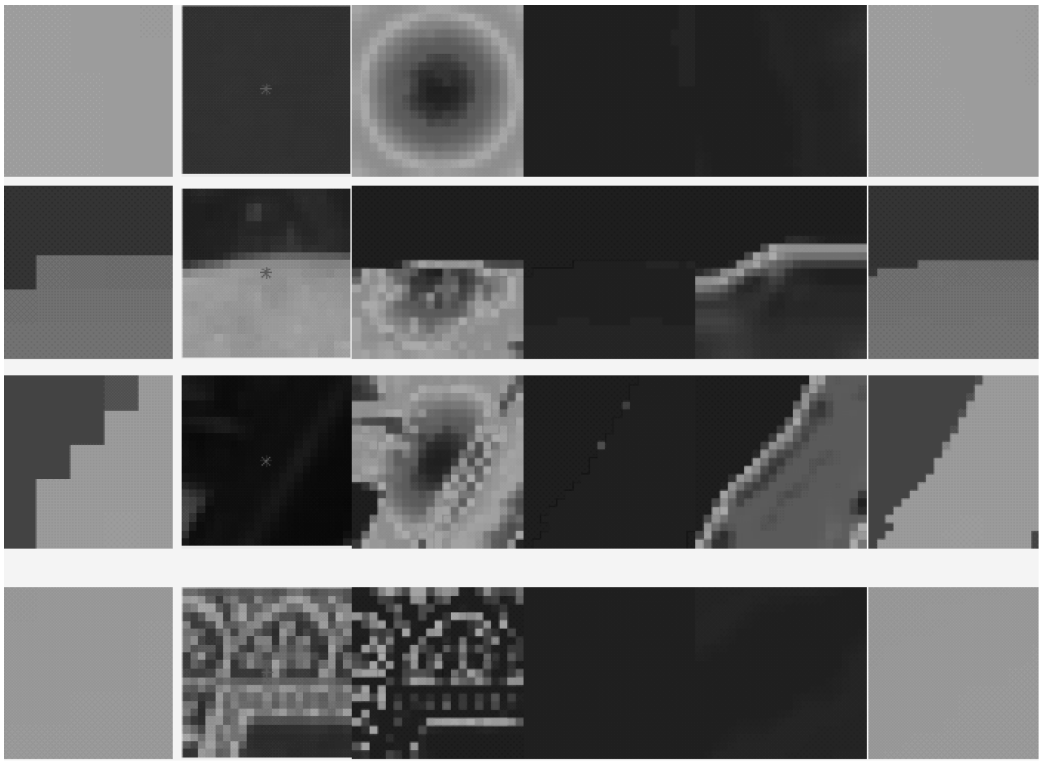


图 1　滤波器核

另一类 DSR 算法需要一个外部数据库。文献[6-8]是其中具有代表性的方法:外部数据中存储着大量低分辨率和高分辨率的深度图像块对,对于输入的待上采样低分辨率深度图像块,此类方法在数据库中做块匹配,从而输出高分辨率图像块。在很多计算机视觉领域的课题中,深度神经网络展现了从海量数据充分挖掘信息的能力。文献[8]针对颜色图像(RGB 图像)超分辨问题,提出 SRCNN 算法,利用深度卷积神经网络学习低/高分辨率图像块对的端到端映射。然而,不同于颜色图,深度图中往往包含着尖锐的边缘,直接将 SRCNN 算法应用于 DSR,将无法保留住这些锐利的边缘。

2　相关研究

2.1　颜色图超分辨率

文献[9]从外部数据库学习低/高分辨率图像块间的单射,通过马尔科夫随机域(Markov Random Field,MRF)将匹配到的高分辨率块重构成结果图。文献[10]提出在原低分辨率图像中挖掘图像块信息冗余以查找高分辨块候选,因此不需要外部数据库。稀疏表示方法[11-16]从一个由低/高分辨率图像块对组成的数据库中学习字典。超分辨率时,对于给定的低分辨率图像块,此类算法将其在低分辨率字典中编码,然后将其表达为对应高分辨率字典中的稀疏线性组合。文献[8]直接利用卷积神经网络学习了一个低/高分辨率图像块的端到端映射,这是图像超分辨率领域首次针对卷积神经网络的应用。

2.2　单深度图超分辨率

2.1 节中提到的图像超分辨率算法可以修改应用到 DSR 问题。在外部数据库中为低分辨率深度图像块搜索候选高分辨率块[6-7]。因此,这实际上是马尔科夫随机域标签(Markov Random Field Labeling)问题。文献[9]提出依次从待超分辨率深度图中不重叠地提取图像块并匹配高分辨率块。与文献[9]不同的是,文献[6]直接在低分辨率空间做块匹配,以此减少模糊现象及匹配复杂度。文献[7]提出一种像素收集算法,作为对文献[6]的改进,它以匹配到的高分辨图像块的中心像素值作为候选而不是整个图像块。而且,当对应颜色图可获得时,文

献[7]还可通过添加能量方程项提升算法表现。

2.3 结合颜色图的深度超分辨率

对于 DSR 问题,场景的颜色图已被证明是一个很有效的信息来源。文献[17]将其建模成一个 MRF 问题并采用共轭梯度算法解决。文献[18]采用保边滤波器如联合双边滤波器[19-20],以颜色图为指导图进行滤波。文献[21]同时提出了基于联合双边滤波器的通用图像上采样算法。此类方法基于这样一个隐含假设:颜色相近的像素具有相近的深度值。然而,这个假设经常会被违背:1)深度图平滑颜色图有纹理时,结果中会出现纹理复制。2)深度图有纹理颜色图平滑时,结果中会出现边缘模糊现象。文献[22]提出了一种噪声敏感滤波器,在深度平滑区域采用深度图像作为指导图,同时在深度不连续区域采用颜色图作为指导图,文献[22]克服了纹理复制现象,但仍然受限于边缘模糊。

2.4 卷积神经网络

文献[23]创建了卷积神经网络(Convolutional Neural Network,CNN)模型。文献[24]在 ImageNet[25]数据集上取得了较大的成果。CNN 模型还有另外一些在计算机视觉领域非常成功的应用:场景解析[26],阴影检测[27],视频分类[28],目标检测[29],图像超分辨率[8],图像恢复[30]。矫正线性单元[31](Rectified Linear Unit,ReLU)是关键设计,它加快了收敛过程并得到了更好的收敛结果。本文的研究也得益于 Matconvnet 工具包[32],它是一个 CNN 模型的高效实现并易于修改的网络结构。

3 本文算法的构造及特点

3.1 基于滤波的 DSR 算法

目前最理想的 DSR 算法是用场景高分辨率颜色图作为指导图进行低分辨率/稀疏深度图的上采样。用 D 和 L 分别表示超分辨率后的结果图和原始低分辨率深度图,q 表示 D 中的像素坐标,$q_{\downarrow}$ 表示 L 中对应像素的坐标。基于滤波的 DSR 算法可被公式化为:

$$D_p = \sum_{q_{\downarrow}\in\Omega_p}(w_{p,q}L_{q_{\downarrow}})/\sum_{q_{\downarrow}\in\Omega_p}w_{p,q} \tag{1}$$

其中,$w_{p,q}$表示像素 p 的滤波器核中,像素 q 的权重。用 G 表示指导图,则联合双边滤波器可以被表示为:

$$w_{p,q}=\exp\left(-\frac{\|p-q\|^2}{2\sigma_S^2}\right)\exp\left(-\frac{|G_p-G_q|^2}{2\sigma_R^2}\right) \tag{2}$$

其中,σ_S 和 σ_R 分别表示 2 个调整空间相似度和范围相似度的常数。图 1(c)展示了以图 1(b)中所示的高分辨率颜色图作为指导图计算出的联合双边滤波器核,其中,$\sigma_S=10$,$\sigma_R=10$。

最佳滤波器核可用场景真实的高分辨深度图作为指导图计算得到。即在最佳滤波器核中,当像素 q 真实深度值与中心像素 p 深度值差异较大时,对应的权重 $w_{p,q}$要足够小。

然而,由于不可避免的设备噪音,得到理论上精准最佳滤波器核在实际应用中是不现实的。因此,做了这样一个近似处理:图 1(d)中所示最佳滤波器核通过将 σ_S,σ_R 分别设置为 $+\infty$和 5 得到。理论上,σ_S/σ_R 应该有足够的大/小来保证最大化/最小化来自正确/错误的种子深度值的贡献。在实际应用中,σ_R 被设置为相对小的值来避免设备噪音,同时 σ_S 被设置为无穷大以忽略空间相似性。

3.2 DSR 算法中的数据驱动滤波器核

如 2.3 节中的讨论,由于不完善的场景颜色信息的使用方式,在基于保边滤波的 DSR 算法中,会发生纹理复制和边缘模糊现象。如图 1 的最后 2 行所示,这些缺陷的发生主要由于:1)2 个物体的深度值不同,但是可能有类似的颜色(如在阴暗的环境中);2)同一个物体内可能有纹理状的深度分布(如树干表面)。因此,对于基于保边滤波的 DSR 算法,适当且高效地使用场景颜色信息至关重要。本文测试了不同种人为设计滤波器核的保边滤波算法,包括联合双边滤波[21,33]、指导图滤波[34]、交叉局部多点滤波(Cross-based Local Multipoint Filtering,CLMF)[35]、树形滤波[36]。

CNN 展现出可以高效地学习有价值的特征,其表现需要很多人为设计。受此启发,本文提出利用 CNN 模型学习针对 DSR 算法的最佳滤波器核。用 H 表示输入低分辨深度图 L 的双三次(Bicubic)上采样结果,$patch_p^H$ 表示从图像 H 上截取的以 p 为中心的图像块。本文的目标是用 CNN 模型学习一个 $patch_p^H$ 到对应局部滤波器核的映射函数:

$$w_p=f_{\mathrm{CNN}}(patch_p^H) \tag{3}$$

以上的模型过程是 CNN 在计算滤波器核中的直接应用。然而,由于颜色信息的缺失,它的表现低于预期。用 G 表示指导图(例如颜色图),一个简单的应对方法是将 $patch_p^H$ 和 $patch_p^G$ 同时输入到网络中,然而在训练过程中发现,由于无法获得大规模高质量的训练数据,这个简单的集成没有收到效果,训练过程无法收敛。

为了降低对训练数据量的依赖,受联合双边滤波过程启发,在输入到网络前,需要将 $patch_p^H$ 减去一个值 d_p。理论上,d_p 应为像素 p 的真实深度值,这个

模型可以表示为:

$$w_p = f_{\mathrm{CNN}}(patch_p^H - d_p) \tag{4}$$

实际中这个真实深度值 d_p 是无法获得的,因此,本文提出了一种 d_p 的估计算法。

d_p 的估计值是从原始低分辨率深度图 L 得到的。用 $\tilde{L}_p$ 表示估计值候选集,$p_\downarrow$ 表示像素 p 在低分辨深度图 L 中的对应像素。在图 L 上,$\tilde{L}_p$ 由像素 $p_\downarrow$ 的邻接像素组成:

$$q_\downarrow \in \tilde{L}_p, \text{if } \| q_\downarrow - p_\downarrow \|^2 \leqslant 2 \tag{5}$$

最后,候选值的颜色距离被用来选择 d_p 的估计值:

$$d_p = L_{\underset{i_\downarrow \in \tilde{L}_p}{\operatorname{argmin}} |G_i - G_p|} \tag{6}$$

其中,$i_\downarrow$ 表示像素 i 在低分辨图 L 中的对应像素。

以上的解决方案用一步预处理的方式将颜色图/指导图信息整合到 CNN 中,同时极大降低了对训练数据规模的要求。将 SRCNN 算法直接应用于 DSR 问题,在非深度边缘区域,结果还是足够理想的。这表明,在深度平滑区域,颜色信息对于滤波器核的贡献应该远小于稀疏深度信息(否则会发生纹理复制现象)。由此,本文提出的模型恰当且高效地融入了颜色信息。

3.3　训练数据

本文使用的训练数据由 Middlebury 2003,2006 全尺寸数据集中的 21 个场景组成[37]。每个场景包括 2 个视角,每个视角包含一张高分辨深度图及其对应的高分辨率颜色图。如 3.1 节中的论述,网络的输出标签由式(2)计算得到,其中,$\sigma_S = +\infty$,$\sigma_R = 5$,指导图为场景高分辨率深度图。待上采样深度图由场景高分辨率深度图通过最近邻插值下采样得到。网络的输入由式(4)、式(6)计算得到。

3.4　网络结构

本节讨论 CNN 模型的具体结构。本文的卷积神经网络有 3 个卷积层,每个卷积层对其输入做卷积并跟随一个激活函数层。每个卷积层包含一个尺寸为 $s_1 \times s_2 \times s_3 \times s_4$ 的卷积核仓库 F 和一个 s_4 维的偏置向量 $\boldsymbol{B}$。其中,s_1、s_4 分别为此层的输入,输出特征图数,$s_2 \times s_3$ 为卷积运算的空间支持域。更形象地说,每个卷积层对输入的 s_1 张特征图做卷积,输出 s_4 张特征图,偏置向量 $\boldsymbol{B}$ 中的每个元素分别绑定到每张输出特征图上进行值偏置。激活层采用 ReLU 函数[31],它可以加快训练过程并得到更好的收敛效果[24]。下文将一个卷积层及其跟随的激活层称为一个块。块 i 可被表达为函数 f_i:

$$f_i(x_i) = x_{i+1} = \max(F_i \times x_i + \boldsymbol{B}_i, 0) \tag{7}$$

其中,x_i 为块 i 的输入,F_i、$\boldsymbol{B}_i$ 分别为块 i 的卷积核仓库和偏置向量。最后,整个网络可被表达为:

$$\begin{aligned} w_p &= f_{\mathrm{CNN}}(patch_p^H - d_p) \\ &= f_3(f_2(f_1(patch_p^H - d_p))) \end{aligned} \tag{8}$$

映射函数 f_{CNN} 的参数可记为 $\theta = \{F_1, \boldsymbol{B}_1, F_2, \boldsymbol{B}_2, F_3, \boldsymbol{B}_3\}$,网络的学习通过最小化网络输出 w_p 和输出标签 w_p^{GT} 的损失函数实现,本文采用均方误差(Mean Squared Error)作为代价函数:

$$\begin{aligned} J(\theta) &= \frac{1}{2n}\sum_p \| w_p - w_p^{GT} \|^2 \\ &= \frac{1}{2n}\sum_p \| f_{\mathrm{CNN}}(patch_p^H - d_p) - w_p^{GT} \|^2 \end{aligned} \tag{9}$$

其中,n 为训练图像块数。

3.5　算法实现

在训练阶段,一个块对 $\{patch_p^H, w_p^{GT}\}$ 在选中的像素 p 处被提取出来。在本文中,$patch_p^H$ 的尺寸是 31×31。3 个卷积层的卷积核仓库尺寸分别是 $1 \times 8 \times 8 \times 64$、$64 \times 3 \times 3 \times 32$、$32 \times 2 \times 2 \times 1$。由于卷积运算特性,卷积层的输出特征图尺寸会小于输入特征图尺寸。根据 3 个卷积层的空间支持域,网络输出标签 w_p^{GT} 的尺寸为 21×21。

所有训练块对都是在深度非连续区域选取的[6-7]。深度边缘通过在场景高分辨率深度图中用 Canny 算子检测得到。检测得到的边缘膨胀 5 个像素来定位深度非连续区域。当且仅当像素 p 位于这个区域时才提取训练块对,并以 6 像素的步长跳跃选择。最终约 40 000 个训练块对被提取出来,训练得到 3 个超分辨率倍数分别为 4、8、16 的 CNN 模型。

损失函数的最优化采用随机梯度下降法(Stochastic Gradient Descent, SGD),在第 t 步,网络参数的更新可被表示为:

$$\theta^t = \theta^{t-1} - r\frac{\partial J(\theta)}{\partial \theta} \tag{10}$$

其中,r 为学习率。本文中学习率被设置为 10^{-5}。卷积层中参数 $\{F_1, F_2, F_3\}$ 采用均值 0,标准差 0.01 的高斯分布初始化。偏置值 $\{\boldsymbol{B}_1, \boldsymbol{B}_2, \boldsymbol{B}_3\}$ 采用常数 0 初始化。

4　实验与结果分析

本文在 Middlebury2005[37] 和 ToFMark[38] 数据

集上进行了实验。Middlebury2005 数据集由 6 个场景组成：Art，Books，Dolls，Laundry，Moebius，Reindeer。ToFMark 数据集由 3 个场景组成：Books，Devil，Shark。每个场景由一张深度图及其对应的颜色图组成。

本文提出的 DSR 算法将在定量评价和定性评价 2 个维度与目前最先进的算法进行比较。这些算法可被分为 2 大类：

1)需要场景颜色图：文献[33]的联合双边滤波(JBFcv)算法，文献[36]的树形滤波(Tree)算法，文献[41]的自适应回归(AP)算法，文献[34]的指导图滤波(Guided)算法，文献[38]的全局广义变分(TGV)算法，文献[40]的联合几何上采样(JGF)算法，文献[39]的边权重规则化(Edge)算法，文献[35]的交叉局部多点滤波(CLMF)算法。

2)单深度图上采样：双三次插值(Bicubic)，文献[6]的块匹配(PB)算法，文献[8]的 SRCNN 算法。

本文测试的算法源代码由原作者提供或者发布。Bicubic 算法、Guided 算法、CLMF 算法并不是专门针对 DSR 问题设计，因此在本文实验中挑选了最佳参数，其他算法的参数使用它们的默认参数和设置。对于基于训练的 PB 算法和 SRCNN 算法，使用了作者训练并发布的模型，因此，下文实验中只比较它们 4 倍超分辨率的结果。与其他 DSR 算法相同，为了维持深度图中锐利的边缘，输入的待超分辨率深度图由原始高分辨率深度图最近邻下采样得到。

4.1 定量评价

表 1 和表 2 用 2 种常用度量标准展示了本文算法与其他算法的数值比较：错误像素率(Percentage of Bad Pixels)和平均绝对差(Mean Absolute Difference，MAD)。当视差误差小于 1 时，一个像素被称为错误像素。令 Bic 算法、文献[35]算法 1、文献[35]算法 2、文献[6]算法、文献[8]算法、文献[38]算法、文献[34]算法、文献[33]算法、文献[39]算法、文献[40]算法、文献[41]算法、文献[36]算法和本文算法分别表示为算法 1 ~ 算法 13。表 1、表 2 中最好的表现加粗表示，本文算法在所有超分辨率倍数上的平均性能优于其他算法。

表 1 Middlebury 数据集上各算法错误像素率比较

数据集	倍数	算法 1	算法 2	算法 3	算法 4	算法 5	算法 6	算法 7	算法 8	算法 9	算法 10	算法 11	算法 12	算法 13
Art	4	10.5	7.57	8.12	3.12	7.61	5.14	9.97	3.36	6.82	3.25	4.13	3.96	**2.28**
	8	19.5	16.7	17.2	—	—	10.5	15.5	8.73	13.49	7.39	5.58	5.24	**4.27**
	16	35.1	33.3	33.2	—	—	21.3	28.4	21.7	25.9	14.3	21.6	9.74	**8.61**
Book	4	3.80	3.17	3.27	**1.39**	2.88	2.48	3.68	4.05	3.35	2.14	1.88	5.77	1.65
	8	8.15	7.25	7.25	—	—	4.65	6.52	10.1	8.55	5.41	4.16	7.22	**3.51**
	16	16.4	16.9	16.1	—	—	11.2	13.0	19.9	19.3	12.0	9.25	11.4	**8.04**
Dolls	4	4.73	3.97	4.04	13.9	3.93	4.45	4.46	3.98	2.90	3.23	4.07	4.60	**2.01**
	8	9.60	9.65	8.76	—	—	11.1	7.63	12.8	6.84	7.29	6.62	6.36	**4.53**
	16	19.5	18.3	18.3	—	—	45.5	15.7	29.7	17.9	15.8	11.5	13.0	**10.9**
Laundry	4	7.42	6.11	5.50	20.6	6.25	6.99	6.33	2.39	2.82	2.60	3.51	2.27	**1.55**
	8	14.8	12.5	12.6	—	—	16.3	11.9	5.64	5.46	4.54	5.19	3.94	**2.71**
	16	26.7	25.3	25.4	—	—	53.6	20.2	13.7	13.5	8.69	11.1	8.87	**7.56**
Moebius	4	4.54	4.03	4.13	**1.95**	3.63	3.68	4.78	3.19	3.72	3.36	2.14	3.52	2.25
	8	9.38	8.40	8.42	—	—	6.84	7.88	7.43	7.36	6.45	5.57	4.90	**3.98**
	16	18.5	17.6	17.2	—	—	14.9	14.8	15.7	14.0	12.3	10.8	8.67	**7.41**
Reindeer	4	5.20	4.60	4.65	6.04	3.84	4.67	5.16	3.89	2.67	2.27	3.64	3.97	**1.59**
	8	9.88	9.71	9.96	—	—	11.2	8.11	13.9	6.22	5.17	5.76	5.76	**2.84**
	16	19.3	18.2	18.3	—	—	43.4	15.7	27.1	16.8	11.8	9.40	12.7	**7.42**
Average	4	6.04	4.91	4.95	7.86	4.69	4.57	5.73	3.48	3.71	2.81	3.23	4.02	**1.89**
	8	11.9	10.7	10.7	—	—	10.11	9.59	9.80	7.99	6.04	5.48	5.57	**3.64**
	16	22.6	21.6	21.4	—	—	31.5	18.0	21.3	17.9	12.5	12.3	10.7	**8.32**

表 2　Middlebury 数据集上各算法 MAD 值比较

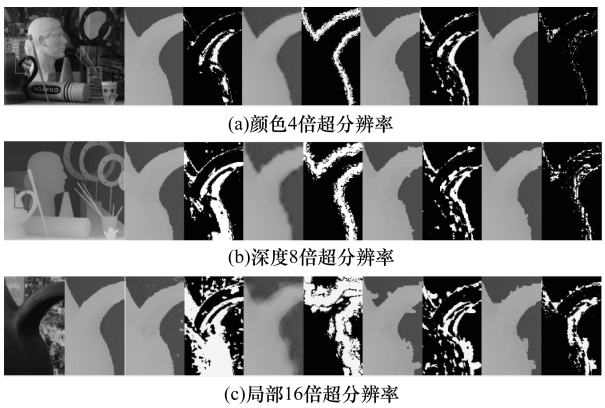
数据集	倍数	算法 1	算法 2	算法 3	算法 4	算法 5	算法 6	算法 7	算法 8	算法 9	算法 10	算法 11	算法 12	算法 13
Art	4	0.97	0.74	0.76	0.93	0.63	0.65	0.96	0.55	0.65	0.47	0.49	0.67	**0.45**
	8	1.85	1.37	1.44	—	—	1.17	1.57	1.08	1.03	0.78	**0.64**	0.84	0.74
	16	3.59	2.95	2.87	—	—	2.30	3.05	2.26	2.11	1.54	2.01	**1.49**	1.55
Book	4	0.29	0.28	0.28	**0.16**	0.25	0.27	0.35	0.38	0.30	0.24	0.22	0.46	0.22
	8	0.59	0.51	0.51	—	—	0.42	0.58	0.71	0.56	0.43	**0.37**	0.55	0.39
	16	1.15	1.06	1.02	—	—	0.82	1.06	1.40	1.03	0.81	0.77	0.84	**0.74**
Dolls	4	0.36	0.34	0.34	0.83	0.29	0.33	0.36	0.41	0.31	0.33	0.34	0.48	**0.27**
	8	0.66	0.66	0.60	—	—	0.70	0.56	0.82	0.56	0.59	0.50	0.58	**0.46**
	16	1.18	1.02	1.01	—	—	2.20	1.01	1.80	1.05	1.06	0.82	0.94	**0.82**
Laundry	4	0.54	0.50	0.50	1.13	0.40	0.55	0.51	0.33	0.32	0.36	0.34	0.41	**0.26**
	8	1.04	0.82	0.80	—	—	1.22	0.89	0.61	0.54	0.64	0.53	0.56	**0.44**
	16	1.95	1.66	1.67	—	—	3.37	1.65	1.33	1.14	1.20	1.12	0.95	**0.94**
Moebius	4	0.30	0.29	0.26	**0.17**	0.25	0.29	0.34	0.33	0.29	0.25	0.20	0.40	0.25
	8	0.59	0.52	0.51	—	—	0.49	0.55	0.68	0.51	0.46	**0.40**	0.49	0.41
	16	1.13	1.01	0.97	—	—	0.90	1.00	1.44	1.10	0.80	0.79	0.82	**0.74**
Reindeer	4	0.55	0.51	0.51	0.56	0.35	0.49	0.54	0.45	0.37	0.38	0.40	0.48	**0.31**
	8	0.99	0.84	0.84	—	—	1.03	0.83	0.90	0.63	0.64	0.58	0.62	**0.48**
	16	1.88	1.51	1.55	—	—	3.05	1.64	1.77	1.28	1.09	1.00	1.04	**0.97**
Average	4	0.50	0.44	0.45	0.54	0.36	0.44	0.51	0.41	0.39	0.37	0.33	0.48	**0.29**
	8	0.95	0.79	0.78	—	—	0.84	0.83	0.80	0.65	0.67	0.50	0.60	**0.49**
	16	1.81	1.52	1.53	—	—	2.10	1.57	1.67	1.25	1.23	1.09	1.01	**0.96**

表 1 比较了错误像素率,衡量了各算法结果中正常值的准确程度。标准超分辨率算法如双三次插值在此表中表现较差。由于不能保持尖锐的深度边缘,它们在深度边缘准确度很低。本文算法只在4 倍上采样的 2 个场景表现略差于 PB[6] 算法,这表明本文算法的上采样准确度非常高(尤其是在深度边缘附近)。这得益于本文提出的数据驱动滤波器核的应用。

表 2 比较了 MAD 值,因而更侧重于评价整体表现。不同于错误像素率这个统计指标,它并不是特别强调维持尖锐的深度边缘。因此,相对于表 1,基于保边滤波的 DSR 算法与传统双三次插值的表现差距大大缩小。如 3.5 节中所述,本文的训练数据仅在深度边缘附近选取,因此主要侧重于尖锐的深度边缘的恢复。尽管如此,本文提出的算法仍然是所有超分辨率倍数上 MAD 值最低的(如表 2 所示,在 67% 的场景中,本文算法超越了所有其他算法)。

4.2　定性评价

图 2 选取了 Middlebury 数据集中的 Arts 场景,将本文算法与流行的基于保边滤波的算法[33-34,36]进行比较。图 2(a)、图 2(b)是颜色图和场景真实高分辨深度(视差)图。图 2(c)是场景中一个颜色和深度边缘不一致区域的特写。图 2 的右半部分分别表示文献[33]算法、文献[34]算法、文献[36]算法、本文算法选择区域的上采样结果图及对应的错误图(以阈值 1 进行二值化得到)。文献[33-34,36]算法主要依靠稠密颜色边缘进行深度边缘上采样,当深度边缘与颜色边缘不一致时算法准确度大大下降。本文提出的 CNN 模型同时利用颜色和深度信息,在深度边缘附近得到了更为准确的结果。



(a)颜色4倍超分辨率

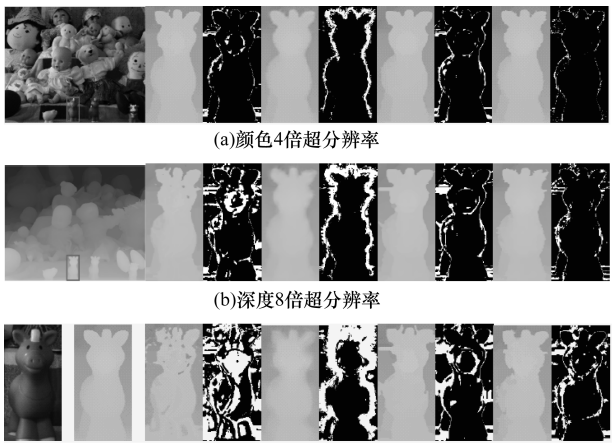
(b)深度8倍超分辨率

(c)局部16倍超分辨率

图 2　深度边缘模糊现象

图 3 用 Middlebury 数据集中的 Dolls 场景,展现了基于保边滤波算法可能会出现的纹理复制现象。图 3(a)、图 3(b)是颜色图和场景真实高分辨深度(视差)图。图 3(c)是场景的选择区域的局部特写。

右半部分本别表示文献[33]算法、文献[34]算法、文献[36]算法、本文算法选择区域的上采样结果图及对应的错误图(以阈值 1 进行二值化得到)。文献[33,36]算法在超分辨倍数较高时有明显的纹理复制现象,例如玩具驴的眼睛和耳朵附近。本文算法能更好地抑制这些现象。



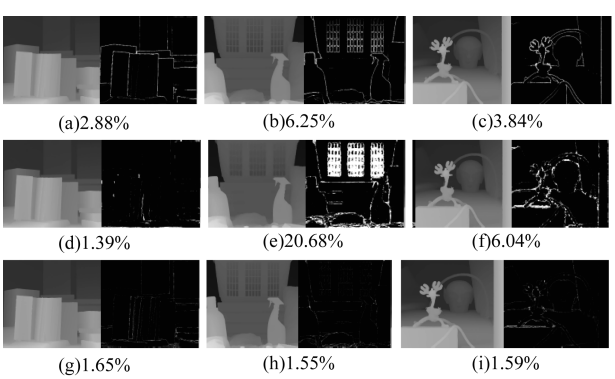
(a)颜色4倍超分辨率

(b)深度8倍超分辨率

(c)局部16倍超分辨率

图 3 纹理复制现象

图 4 用 Middlebury 数据集中的 Book,Laundry,Reindeer 场景比较了基于学习的 DSR 算法。其中,图 4(a)、图 4(c)依次为 SRCNN 算法、PB 算法、本文算法的超分辨结果及对应错误图,超分辨倍数为 4 倍。图下的数字为对应的错误像素率。SRCNN 主要针对颜色图上采样设计,区别于深度图,颜色图中尖锐的边缘非常罕见。因此,如图 4(a)所示,SRCNN 算法趋向于模糊掉尖锐的深度边缘。如图 4(b)所示,PB 算法成功恢复出了非常锐利的深度边缘,但是在细微结构处表现较差。这是因为,PB 算法主要依赖输入深度图与训练数据的场景相似性获得高分辨率输出。在低分辨率空间,这些细微结构会丢失细节。本文算法将场景颜色信息融入模型,在结果图中得到了更高质量的深度边缘。



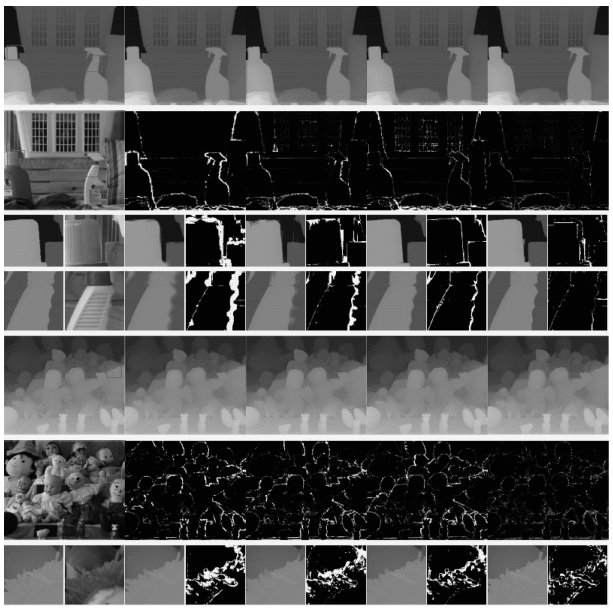
(a)2.88% (b)6.25% (c)3.84%

(d)1.39% (e)20.68% (f)6.04%

(g)1.65% (h)1.55% (i)1.59%

图 4 训练算法结果展示

图 5 选取了 Middlebury 数据集中的 Laundry,Dolls 场景来全面比较本文算法和 3 个在表 1、表 2 中数值表现比较好的算法。从图 5(b)、图 5(c)的局部图可以发现,AP 算法[41]可能会模糊深度边缘,同时 JGF 算法[40]会产生错误边缘。Edge 算法[39]的深度边缘更好一些。然而类似于文献[40-41]算法,它也会在具有相似颜色的像素间丢失深度细节(如图 5(d)的最后一行所示)。如图 5(e)所示,本文算法针对颜色深度状态不一致情况下有更好的鲁棒性。



(a)场景 (b)AP算法 (c)JGF算法 (d)Edge算法 (e)本文算法

图 5 8 倍超分辨率结果比较

除了 Middlebury 数据集,本文也在 ToFMark[38]数据集上做了定量分析,如表 3 所示。表 3 展示的是 4 倍超分辨结果,用 MAD 作为度量标准,本文算法在所有场景中都取得了最小的 MAD 值。

表 3 ToFMark 数据集上各算法错误像素率比较

算法	Books	Devil	Shark	Average
Tree 算法	5.34	15.07	13.72	11.38
PB 算法	3.91	3.44	5.37	4.24
SRCNN 算法	3.59	2.59	4.48	3.56
Guided 算法	3.50	2.47	4.57	3.51
JBFcv 算法	3.38	2.85	3.95	3.39
TGV 算法	3.19	2.44	4.07	3.23
本文算法	**2.54**	**2.07**	**3.29**	**2.63**

4.3 与 SRCNN 算法的比较

本节继续深入比较本文算法与 SRCNN 算法[8]。原始 SRCNN 模型使用颜色图训练得到,并不适合于直接应用于 DSR 问题。因此本文特别用深度图重新训练了一个 SRCNN 模型。然而如图 6

所示,SRCNN 模型并不适合直接用深度图训练。深度图与颜色图非常不同,深度图中纹理少,含有尖锐且稀疏的边缘。当用深度图训练时,神经网络倾向于在训练过程中忽略这些边缘。深度图训练的 SRCNN 表现要比原始模型性能更差。

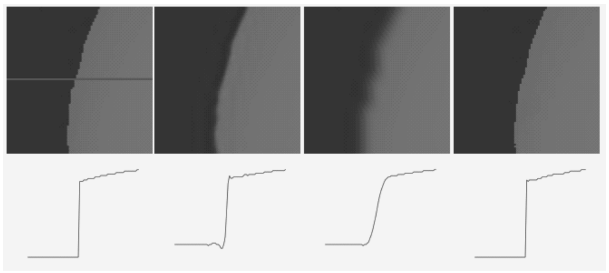


图 6　本文算法与 SRCNN 算法比较

如前文的讨论及图 2 和图 4(a)所示,SRCNN 会模糊掉深度边缘。然而,维持尖锐的边缘是 DSR 问题中非常重要的方面。当在 3D 中展示时,模糊的深度边缘会产生严重的锯齿化现象。表 1 清晰地展示了本文算法相对于 SRCNN 算法的改进。然而这种改进在表 2、表 3 中有了一定程度的缩小,这主要是由于深度图中的边缘是非常稀疏的(尤其在 ToFMark 数据集中)。因此,MAD 值并不能完全展示本文算法相对于 SRCNN 算法的改进。尽管如此,本文算法的平均 MAD 值在 Middlebury 和 ToFMark 数据集上的所有场景仍然是最低的。

5　结束语

受 CNN 在图像超分辨领域成功的启发[8],本文提出在场景颜色图辅助下应用 CNN 进行深度图超分辨。保边滤波器已被证明对 DSR 问题非常有效。但是此类算法主要目的是将颜色信息迁移到超分辨率结果图中,当颜色和深度纹理/边缘不匹配时,结果会发生很明显的错误。为了更恰当地同时利用稠密/高分辨率颜色信息和稀疏/低分辨率深度信息,提出应用 CNN 学习基于数据驱动的保边滤波器核。大量实验结果证明了本文提出模型的高效性。与其他基于滤波或者学习的算法相似,本文算法没有考虑含有噪声的深度输入及颜色指导图,这是今后需要研究的方向。

参考文献

[1] KUSTER C, POPA T, ZACH C, et al. FreeCam: A Hybrid Camera System for Interactive Free-viewpoint [C]// Proceedings of Video Vision, Modeling & Visualization Workshop. Berlin, Germany: Springer, 2011: 17-24.

[2] KIMY M, THEOBALT C, DIEBEL J, et al. Multi-view Image and ToFSensor Fusion for Dense 3D Reconstruction[C]// Proceedings of the 12th International Conference on Computer Vision. Washington D. C., USA: IEEE Press, 2009: 1542-1549.

[3] HOLZ D, SCHNABEL R, DROESCHEL D, et al. Towards Semantic Scene Analysis with Time-of-Flight Cameras[C]//Proceedings of Conference on Robot Soccer World Cup XIV. Berlin, Germany: Springer, 2011: 121-132.

[4] BOHME M, HAKER M, MARTINETZ T, et al. A Facial Feature Tracker for Human-computer Interaction Based on 3D ToF Cameras [J]. International Journal of Intelligent Systems Technologies and Applications, 2008, 5(3): 264-273.

[5] SHOTTON J, SHARP T, KIPMAN A, et al. Real-time Human Pose Recognition in Parts from Single Depth Images[J]. Computer Vision & Pattern Recognition, 2011, 56(1): 116-124.

[6] AODHA O M, CAMPBELL N D, NAIR A, et al. Patch Based Synthesis for Single Depth Image Super-resolution[C]// Proceedings of European Conference on Computer Vision. Berlin, Germany: Springer, 2012: 71-84.

[7] LI Jing, LU Zhichao, ZENG Gang, et al. Similarity-aware Patchwork Assembly for Depth Image Super-resolution[C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2014: 3374-3381.

[8] CHAO Dong, LOYC C, HE Kaiming, et al. Learning a Deep Convolutional Network for Image Super-resolution[C]//Proceedings of European Conference on Computer Vision. Berlin, Germany: Springer, 2014: 184-199.

[9] FREEMAN W T, PASZTOR E C, CARMICHAEL O T. Learning Low-level Vision[J]. International Journal of Computer Vision, 2000, 40(1): 25-47.

[10] GLASNER D, BAGON S, IRANI M. Super-resolution from a Single Image [C]//Proceedings of the 12th International Conference on Computer Vision. Washington D. C., USA: IEEE Press, 2009: 349-356.

[11] TIMOFTE R, DE V, GOOL L V. Anchored Neighborhood Regression for Fast Example-based Super-resolution [C]//Proceedings of IEEE International Conference on Computer Vision. Washington D. C., USA: IEEE Press, 2013: 1920-1927.

[12] YANG Jianchao, WANG Zhaowen, LIN Zhe, et al. Coupled Dictionary Training for Image Super-resolution[J]. IEEE Transactions on Image Processing, 2012, 21(8): 3467-3478.

[13] YANG Jianchao, WRIGHT J, HUANG T, et al. Image Super-resolution as Sparse Representation of Raw Image Patches[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2008: 1-8.

[14] YANG Jianchao, WRIGHT J, HUANG T S, et al. Image Super-resolution via Sparse Representation [J]. IEEE Transactions on Image Processing, 2010, 19 (11): 2861-2873.

[15] 曹 翔,陈秀宏,潘荣华. 基于稀疏表示的快速图像超分辨率算法[J]. 计算机工程,2015,41(6):211-215.
[16] 练秋生,张 伟. 基于图像块分类稀疏表示的超分辨率重构算法[J]. 电子学报,2012,40(5):920-925.
[17] DIEBEL J, THRUN S. An Application of Markov Random Fields to Range Sensing[C]//Proceedings of Conference on Neural Information Processing Systems. Berlin, Germany: Springer, 2005: 291-298.
[18] 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):40-45.
[19] PETSCHNIGG G, SZELISKI R, AGRAWALA M, et al. Digital Photography with Flash and No-flash Image Pairs[J]. ACM Transactions on Graphics, 2004, 23(3): 664-672.
[20] TOMASI C, MANDUCHI R. Bilateral Filtering for Gray and Color Images[C]//Proceedings of the 6th International Conference on Computer Vision. Washington D. C., USA: IEEE Press, 1998: 839-846.
[21] KOPF J, COHEN M F, LISCHINSKI D, et al. Joint Bilateral Upsampling[J]. ACM Transactions on Graphics, 2007, 26(3): 96.
[22] 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. Berlin, Germany: Springer, 2008: 1-12.
[23] LECUN Y, BOSER B, DENKERJ S, et al. Backpropagation Applied to Handwritten Zip Code Recognition. Neural Computation, 1989, 1(4): 541-551.
[24] KRIZHEVSKY A, SUTSKEVER I, HINTONG E. Image Net Classification with Deep Convolutional Neural Networks[C]//Proceedings of the 25th International Conference on Neural Information Processing Systems. New York, USA: ACM Press, 2012: 1097-1105.
[25] JIA Deng, WEI Dong, SOCHER R, et al. ImageNet: A Large-scale Hierarchical Image Database[C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2009: 248-255.
[26] FARABET C, COUPRIE C, NAJMAN L, et al. Learning Hierarchical Features for Scene Labeling[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35(8): 1915-1929.
[27] KHAN S H, BENNAMOUN M, SOHEL F, et al. Automatic Feature Learning for Robust Shadow Detection[C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2014: 1939-1946.
[28] KARPATHY A, TODERICI G, SHETTY S, et al. Large-scale Video Classification with Convolutional Neural Networks[C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2014: 1725-1732.
[29] SZEGEDY C, TOSHEV A, ERHAN D. Deep Neural Networks for Object Detection[C]//Proceedings of Conference on Neural Information Processing Systems. Berlin, Germany: Springer, 2013: 2553-2561.
[30] EIGEN D, KRISHNAN D, FERGUS R. Restoring an Image Taken Through a Window Covered with Dirt or Rain[C]//Proceedings of IEEE International Conference on Computer Vision. Washington D. C., USA: IEEE Press, 2013: 633-640.
[31] NAIR V, HINTONG E. Rectified Linear Units Improve Restricted Boltzmann Machines[C]//Proceedings of the 27th International Conference on Machine Learning. Washington D. C., USA: IEEE Press, 2010: 807-814.
[32] VEDALDI A, LENC K. Matconvnet: Convolutional Neural Networks for Matlab[C]//Proceeding of ACM International Conference on Multimedia. New York, USA: ACM Press, 2015: 689-692.
[33] YANG Qingxiong, YANG Ruigang, DAVIS J, et al. Spatial-depth Super Resolution for Range Images[C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2007: 1-8.
[34] HE Kaiming, SUN Jian, TANG Xiaoou. Guided Image Filtering[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35(6): 1397-1409.
[35] LU Jiangbo, SHI Keyang, MIN Dongbo, et al. Cross-based Local Multipoint Filtering[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2012: 430-437.
[36] YANG Qingxiong. Stereo Matching Using Tree Filtering[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(4): 834-846.
[37] SCHARSTEIN D, SZELISKI R. Middlebury Stereo Evaluation[EB/OL]. (2015-06-11). http://vision.middlebury.edu/stereo.
[38] FERSTL D, REINBACHER C, RANFTL R, et al. Image Guided Depth Upsampling Using Anisotropic Total Generalized Variation[C]//Proceedings of IEEE International Conference on Computer Vision. Washington D. C., USA: IEEE Press, 2013: 993-1000.
[39] PARK J, KIM H, TAIY W, et al. High Quality Depth Map Upsampling for 3D-ToF Cameras[C]// Proceedings of IEEE International Conference on Computer Vision. Washington D. C., USA: IEEE Press, 2011: 1623-1630.
[40] LIU Mingyu, TUZEL O, TAGUCHI Y. Joint Geodesic Upsampling of Depth Images[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2013: 169-176.
[41] YANG Jingyu, YE Xinchen, LI Kun, et al. Color-guided Depth Recovery from RGB-DData Using an Adaptive Autoregressive Model[J]. IEEE Transactions on Image Processing, 2014, 23(8): 3443-3458.

编辑 顾逸斐