

工作总结与计划

一、工作总结（按时间划分）

日期	工作内容			完成情况及主要问题
	上午	下午	晚上	
星期一	修改开题报告	学 pytorch	读论文	修改开题报告，阅读 pytorch 官方文档
星期二	修改开题报告	读论文	帮舍友做毕设	阅读基于教师-学生网络的去雾网络
星期三	读代码	读论文	读论文	阅读 TIP 超分论文的代码 阅读语义指导下的深度估计表示增强论文
星期四	修改开题报告	调试代码	休息	调试运行 TIP 论文的代码
星期五	读论文	写周报	写周报	阅读深度超分的论文
星期六	做 PPT	做 PPT	做 PPT	提交开题报告，准备开题答辩

二、下一步计划（按任务划分）

编号	工作内容	目标	相关配合
1	总结阅读过的论文	1 月 8 日前完成	无
2	调试深度估计的代码	待定	获得代码
3	准备开题答辩	1 月 8 日前完成	无
4			
5			
主要风险	无		

三、个人分析与总结

内容提要	
1	进度方面：本周总体效率较高，能额外完成计划的任务
2	课题方面：开始代码调试和验证，与老师交流了下一步的规划
3	其他思考：要做好研究计划，加快研究进度
4	

本周主要阅读了三篇论文，深度估计和深度图像超分辨率重建方向各一篇，还有一篇为基于教师-学生网络的图像去雾方向，其中深度估计和图像去雾方向两篇论文设计的网络模型均为“多任务”，从而提高其中一个任务（如深度估计、图像去雾）的性能，而深度图像超分方向则主要将深度学习（即数据驱动）用于拟合出理想的过滤器（filter），总体而言是基于滤波的深度图像超分辨率重建。此外，本周还调试了 2020-TIP 的深度图像超分的代码。

结合上周阅读的深度估计方向的论文来看，当多任务联合学习时，学习过程中的中间特征可以用于损失约束等，可能有助于性能的提高。

四、论文总结

论文标题	Distilling Image Dehazing with Heterogeneous Task Imitation
作者及单位	Ming Hong ¹ , Yuan Xie ^{2*} , Cuihua Li ¹ , Yangyun Qu ¹ ¹ Fujian Key Laboratory of Sensing and Computation for smart City, School of Informatics, Xiamen University, Fujian, China ² School of Computer Science and Technolog, East China Normal Unversity, Shanghai, China
论文出处	2020-CVPR
创新点提炼	1. 提出一个用于图像去雾的知识蒸馏方法，其中教师网络是图像重建的自编码网络，用以辅助图像去雾网络的训练。这是首个用知识蒸馏实现的多任务的图像重建。 2. 提出面向过程的学习机制，面向过程的损失函数，频繁使用的面向结果的保真损失和感知损失一起构成了总的损失函数来度量教师和学生之间的相似性。 3. 为去雾网络设计了一个空间加权的通道注意力残差块和雾密度感知模拟损失。
个人想法	

论文方法及结论：

目前为止，知识蒸馏主要被用于图像分类，语义分割和物体监测，而很难在图像重建上得以实现。这篇论文调研了如何在图像去雾任务上实现知识蒸馏。主要有以下三个难点：

- 用于哪种任务的网络可以在去雾网络的训练中提供有效的知识？** 目前绝大多数知识蒸馏的方法主要使用教师和学生模型来处理相同的任务，例如两个网络都被用于图片分类等。几乎没有调查研究是关于两个不同任务的网络之间的知识迁移，即**是否可以通过不同的任务来帮助训练去雾模型？**
- 教师网络如何在训练的过程中帮助学生网络？** 大部分知识蒸馏方法聚焦于面向结果的学习而忽视了面向过程的学习。例如，Hinton 等人的工作只度量教师与学生网络的输出类别分布的相似性以控制两个网络的学习而忽略了学习过程中丰富的暗知识。**问题在于如何利用学习过程中的暗知识来帮助训练学生网络？**
- 如何度量教师和学生之间的相似性？** 在图像分类的知识蒸馏中，类别分布被用来度量教师和学生网络之间的相似性。然而，对于图像重建而言，没有类别分布可以被用来进行这样的度量。

为了解决上述问题，本文提出了知识蒸馏去雾网络（KDDN）。

对于第一个问题，使用图像重建网络来帮助去雾网络的训练，自编码网络作为教师。

对于第二个问题，本文采用了面向过程的学习机制用于知识迁移。作者监督中间的特征并使用特征的相似性来控制作为学生的去雾网络从作为教师的图像重建网络进行模拟学习（模拟特征）。面向过程的监督充分利用了暗知识从而达到了更好的图像去雾结果。

对于第三个问题，本文将图像的保真损失函数、感知损失函数和教师与学生的中间特征图的像素级差异图结合起来。

论文标题	Semantic-Guided Representation Enhancement for Self-supervised Monocular Trained Depth Estimation
作者及单位	Rui Li, Qing Mao, Pei Wang, Xiantuo He, Yu Zhu, Jinqiu Sun , Yanning Zhang Northwestern Polytechnical University
论文出处	2020-arXiv e-prints
创新点提炼	1. 提出了一个语义引导的边缘强化模块（SEEM），该模块是专门设计用于加强局部的位于语义边缘的基于点的深度表示。 2. 提出了一个语义引导的多尺度注意力模块用于改善全局的深度表示。
算法步骤	

论文方法及结论：

1. 论文提出的问题

当仅给定图像序列作为输入时，自监督的深度估计在生成高质量的深度图方面已经展现出其高效的性能。然而，当估计边缘区域或具有精细结构的物体时，由于其有限的深度表示能力，其性能通常降低。

2. 解决的方法

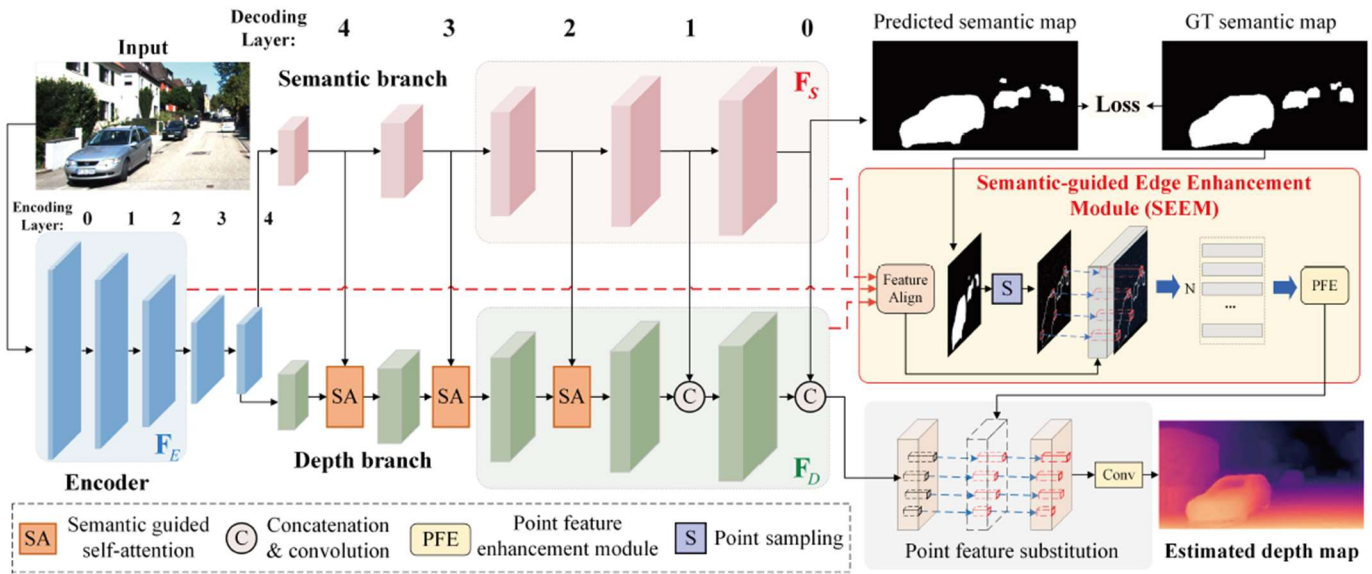


Figure 1: **The overview of the proposed architecture.** An extra semantic branch is proposed to offer contextual information for depth representations. Then, we enhance the local depth feature representation via sampling and enhancing semantic edge-based point features via the proposed *Semantic-guided Edge Enhancement module (SEEM)*. The global feature representation is promoted by the proposed semantic-guided multi-layer self-attention (SA), which explores pixels feature correlations between fused depth and semantic features.

为解决上述问题，本文提出了一个语义引导的深度表示增强方法，可通过丰富的上下文信息促进局部和全局的深度特征表示。与传统范例中使用的单一深度网络不同，本文提出了额外的语义分割分支来为深度估计提供额外的上下文特征。

基于该框架，本文通过采样和将位于语义边缘上的基于点的特征馈送到单个语义引导的边缘增强模块(SEEM)来增强局部特征表示，该模块专门设计用于促进在具有挑战性的语义边界上的深度估计。然后，本文通过提出一种语义引导的多级注意机制来改进全局特征表示，该机制通过探索多级深度解码方案中的像素相关性来增强语义

和深度特征。

语义分支与深度网络共享相同的编码器并一起训练，以获得更好的场景表示。在训练阶段，语义分支以多层的方式将上下文特征馈送到深度分支以加强深度理解。

2.1 SEMANTIC-GUIDED EDGE ENHANCEMENT MODULE (SEEM)

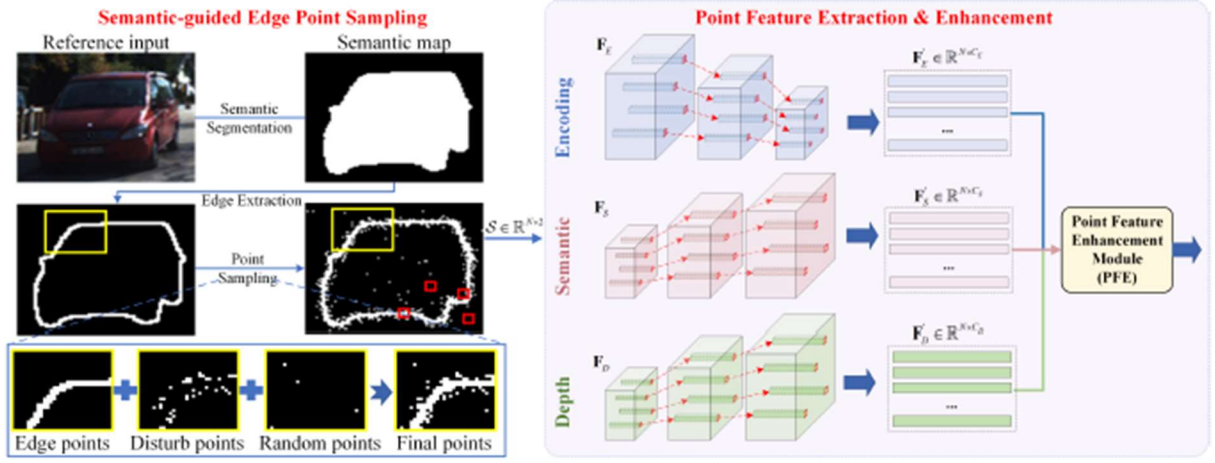


Figure 2: The semantic-guided edge enhancement module (SEEM). The proposed method first samples and extracts the semantic edge-located point set which consists of the edge points, disturbed edge points and random points. Then, the points features are extracted via bilinear sampling and fed into the point feature enhancement module which includes a set of 1-D convolutions, to enhance the local point depth feature representations independently.

将编码特征块 F_E 、深度解码特征块 F_D 以及语义解码特征块 F_S 作为输入，并输出一组点特征 $F' \in R^{N \times C}$ ，其中 N 是采样点数， C 表示与最终输出深度特征一致的输出特征通道。

首先根据语义边缘对局部点进行采样。然后，我们通过网格采样从输入特征块中提取点特征，并通过合并来增强这些特征表示，并将它们馈送到点特征增强模块。产生的代表点特征用于替换相应位置上的最终输出深度特征，以进一步促进语义类别边缘上的深度表示。

Semantic-guided edge point sampling strategy

- (1) 通过卷积二元语义掩码和 sobel 算子计算语义边缘响应
- (2) 提取边点集 $\mathcal{S}_E \in R^{\mu N \times 2}$ ，显示边缘响应图中最大的响应
- (3) 除了正好位于边缘上的点之外，在小范围内选择另一组 μN 均匀采样随机点 $[-c; c]$ ，用这个随机点集去扰动 \mathcal{S}_E 的点位置，得到一个新的扰动边点集 $\mathcal{S}_D \in R^{\mu N \times 2}$
- (4) 在全局图像上随机采样来概括网络的学习能力，生成点集 $\mathcal{S}_R \in R^{(1-\mu N) \times 2}$

最终的点集 $\mathcal{S} \in R^n$ ， $\mathcal{S} = \mathcal{S}_E \cup \mathcal{S}_D \cup \mathcal{S}_R$

Point feature extraction and enhancement

对于的输入特征块 F_E ， F_D ， F_S ，它们可以自由地包含特征层。本文选择层 (2,1,0)，它们被认为为图像、深度和语义表示提供了丰富的信息。我们通过点集 \mathcal{S} 提取点特征 $F'_j = F_j(\mathcal{S}), j \in \{E, D, S\}$ 。然后，将这些来自不同域 (例如，图像、几何和语义域) 的点状特征表示馈送到点特征增强(PFE)模块，该模块首先以通道方式连接输入特征，然后通过一组核大小为 1 的一维卷积来增强点状表示。输出是点特征块 $F'_{out} \in R^{N \times C}$ ， $F'_{out} = f_{\theta}(F'_E, F'_D, F'_S)$ 。用增强的点特征表示 F'_{out} 替换在 \mathcal{S} 对应的位置上深度特征。

2.2 SEMANTIC-GUIDED MULTI-LEVEL SELF-ATTENTION

在本文的方法中，语义分割分支在深度分支的每个解码层提供上下文信息。考虑到深度和语义表示之间的内在相关性，本文提出了多级自注意，其探索融合的深度和语义特征内部的像素相关性，以进一步改进全局特征表示，从而更好地进行深度估计。

注意力集中在深度和语义特征上。在深度解码阶段，假设 F_D^i 和 F_S^i 是深度分支和语义分支第 i 层的特征图，我们连接这两个特征，并通过一组卷积将它们融合，得到特征图 F_A^i 。对于一个给定的融合特征 $F_A^i \in R^{C_i \times H_i \times W_i}$ ，自注意力模块通过卷积操作计算 query, key 和 value， $\{Q_i, Y_i, V_i\} \in R^{C_i \times L_i}$ 。

$$A_i = \text{softmax}(Q_i \cdot Y_i^T)$$

$$F_{\text{out}}^i = \gamma(A_i \cdot V_i) + F_A^i$$

3. 实验结果结论

Table 1: Quantitative results on KITTI 2015. The best results are in **bold** and the second best results are underlined. ‘S’ and ‘M’ refer to self-supervision methods using stereo images and monocular images, respectively. ‘Inst’ and ‘Sem’ mean methods that leverage instance or semantic segmentation information. ‘PN’ and ‘R50’ refer to the method that uses PackNet (Guizilini et al., 2020a) and Resnet-50 as backbone, respectively. All monocular trained methods are reported without post-processing steps. The metrics marked in blue mean ‘lower is better’, while these in red refer to ‘higher is better’. Our method outperforms the state-of-the-arts in most metrics by a large margin.

Method	Training	Abs Rel	Sq Rel	RMSE	RMSE _{log}	$\delta < 1.25$	$\delta < 1.25^2$	$\delta < 1.25^2$
Godard et al. (2017)	S	0.133	1.142	5.533	0.230	0.830	0.936	0.970
Pillai et al. (2019)	S	0.112	0.875	4.958	0.207	0.852	0.947	0.977
Watson et al. (2019)	S	<u>0.106</u>	<u>0.780</u>	4.695	0.193	0.875	0.958	0.980
Godard et al. (2019)	MS	<u>0.106</u>	0.806	4.630	0.193	0.876	0.958	0.980
Zhou et al. (2017)	M	0.183	1.595	6.709	0.270	0.734	0.902	0.959
Mahjourian et al. (2018)	M	0.163	1.240	6.220	0.250	0.762	0.916	0.968
Yin & Shi (2018)	M	0.155	1.296	5.857	0.233	0.793	0.931	0.973
Wang et al. (2018)	M	0.151	1.257	5.583	0.228	0.810	0.936	0.974
Ranjan et al. (2019)	M	0.140	1.070	5.326	0.217	0.826	0.941	0.975
Chen et al. (2019b)	M	0.135	1.070	5.230	0.210	0.841	0.948	0.980
Godard et al. (2019)	M	0.115	0.903	4.863	0.193	0.877	0.959	0.981
Guizilini et al. (2020a)	M	0.111	0.785	<u>4.601</u>	0.189	0.878	0.960	<u>0.982</u>
Johnston & Carneiro (2020)	M	<u>0.106</u>	0.861	4.699	<u>0.185</u>	<u>0.889</u>	0.962	<u>0.982</u>
Casser et al. (2019)	M+Inst	0.141	1.026	5.291	0.215	0.816	0.945	0.979
Chen et al. (2019a)	M+Sem	0.118	0.905	5.096	0.211	0.839	0.945	0.977
Guizilini et al. (2020b)-(PN)	M+Sem	0.102	0.698	4.381	0.178	0.896	0.964	0.984
Guizilini et al. (2020b)-(R50)	M+Sem	0.113	0.831	4.663	0.189	0.878	0.971	0.983
Ours	M+Sem	0.105	0.764	4.550	0.181	0.891	<u>0.965</u>	0.983

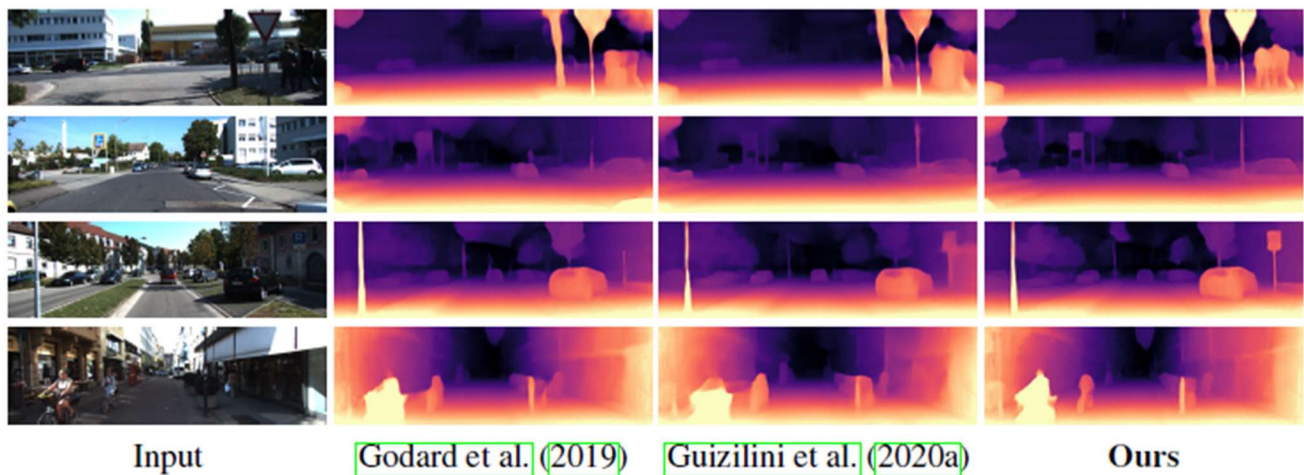


Figure 3: **Qualitative comparison on KITTI 2015.** Our method shows its distinct advantages against other methods in generating sharp depth borders as well as thin structures that is well-aligned with the semantic contextual information.

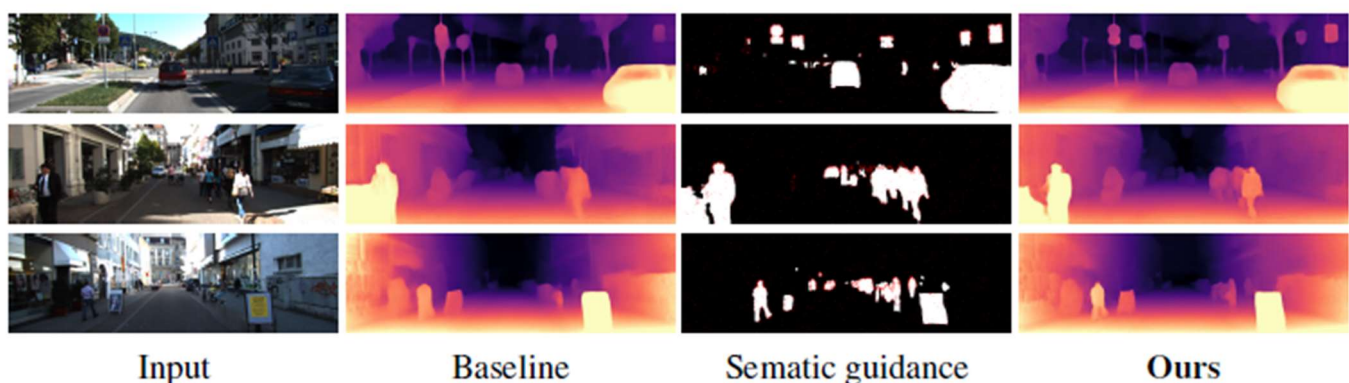


Figure 4: **Qualitative results of the proposed method against the baseline.** The third column refers to the semantic guidance, in which the black and white area denote the foreground and background, and the red dots are the sampled points of the proposed SEEM module.

4. 存在的问题

暂无

论文标题	Deep Color Guided Coarse-to-Fine Convolutional Network Cascade for Depth Image Super-Resolution
作者及单位	Yang Wen ¹ , Bin Sheng ¹ , Ping Li ² , Weiyao Lin ¹ , and David Dagan Feng ³ ¹ Shanghai Jiao Tong University ² Macau University of Science and Technology ³ University of Sydney
论文出处	2019-TIP
创新点提炼	<ol style="list-style-type: none"> 1. 提出一种数据驱动的滤波方法，代替手工设计的滤波器，逼近深度图像超分辨率的理想滤波器。基于大数据样本，学习的滤波器对于深度图像的上采样更准确和稳定。 2. 引入一个由粗到细的 CNN 学习不同大小的过滤器内核。在粗略阶段，CNN 学习较大的滤波器核，以获得粗略的高分辨率深度图像。在精细阶段，使用粗略的高分辨率深度图像作为输入，以便学习更小的滤波器核来获得更精确的结果。 3. 构建融合色差和空间距离的颜色引导策略，用于深度图像上采样。根据高分辨率彩色图中对应的像素点，对插值后的高分辨率深度图像进行修正。在颜色信息的引导下，

	获得的高分辨率图像的深度可以减轻纹理复制伪影，有效保留边缘细节。
算法步骤	

论文方法及结论：

深度图超分辨率重建方法大致可分为基于学习和基于过滤的方法，基于学习的方法由于其较高的计算复杂度而在应用中受到限制。由于颜色信息对于深度图像 SR 提供边缘引导是至关重要的，所以本文应用高分辨率彩色图像作为支持信息，使得具有不同深度的像素可以在上采样过程中根据颜色值被不同地加权。鉴于 CNN 在深度图上采样方面的出色表现，它现在正在逐步结合彩色图像来解决 DSR 问题。由于传统的基于 DSR 滤波的方法不能有效地恢复高频细节，本文使用级联的 CNN 网络和附加的 HR 彩色图像来训练理想的上采样滤波器。

本文提出的模型主要包括三个组成部分:色彩引导预处理，DSR 粗阶段，DSR 精阶段。在下图的颜色引导部分中，首先将输入的低分辨率深度图插值为初始高分辨率版本。由于彩色图像的高频成分如边缘有利于辅助深度像

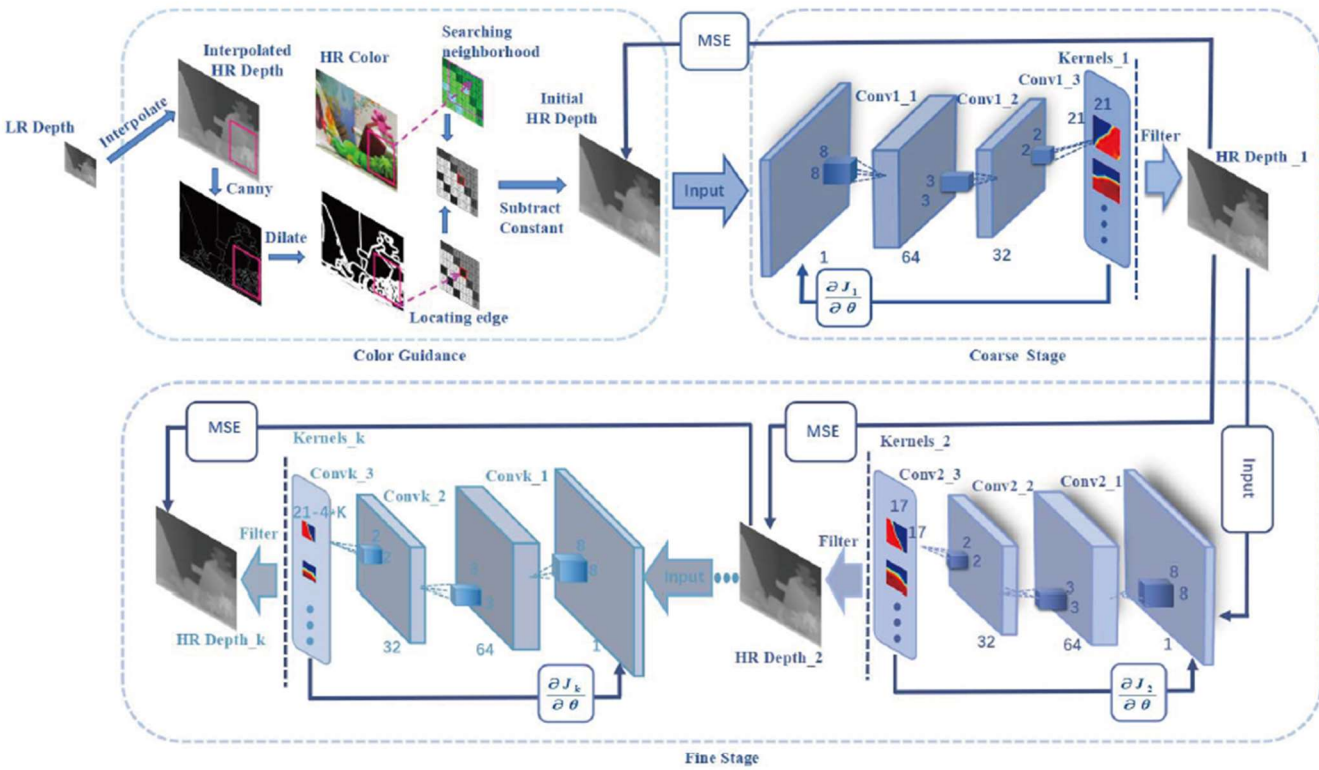


Fig. 2. Conceptual illustration of our framework. It mainly includes three parts: color guidance, coarse stage, and fine stage. In the color guidance part, we first interpolate the LR depth image by bicubic algorithm to obtain the initial HR depth image, then we detect the edges of the interpolated high-resolution depth map by canny operator and dilate the edges by five pixels to determine the marginal areas. To calculate the approximate depth value, both the neighborhood information determined by HR color image and the depth information of LR depth image are utilized. Finally, we subtract the approximate depth from the marginal area of the interpolated high-resolution image. In the coarse stage, the modified interpolated HR depth map is used as the input of CNN to generate the approximate ideal filter, which is named Kernels_1 in the figure. Then the coarse HR depth_1 map can be constructed by filter_1. In order to recover better details, the coarse HR depth_1 is used as input for the fine CNN. Through the fine CNN, smaller kernels are obtained to consider smaller neighborhood and recover better resolution HR depth_k. MSE is used to judge the convergence of network, the network can be considered convergent if MSE is small enough.

素预测，我们将修改边缘区域不确定的插值深度值，以解决边缘不连续和纹理复制问题。

首先通过 Canny 算子得到边缘，然后将其放大 5 个像素。根据彩色图像中心像素与相邻像素之间的最小差值计算局部真实深度常数。最后，基于残差学习思想，从初始高分辨率深度版本中减去局部真实深度常数。对于粗糙阶段，由于基于学习的方法在抑制伪影的同时从外部数据中生成看似合理的细节是非常强大的，所以我们通过三层 CNN 模型学习更大的滤波器核，然后使用这些核来重建名为 **HR Depth_1** 的高分辨率深度图像。如图所示，精细阶段倾向于基于 **HR Depth_1** 生成更多精细细节，因为较小的局部区域可以保留更多的形状边缘。由于 **HR Depth_1** 的分辨率比原始版本高，如果作为精细网络输入，可以通过精细网络获得更高的深度图，同时可以利用不

同 HR 深度图像之间的均方误差来判断网络是否收敛。

Filtering-Based Depth Super Resolution

在附加的高分辨率彩色图像的指导下，基于滤波器的方法对于对低分辨率深度图像进行上采样具有显著的性能。本文引入了上采样理想滤波器的概念，设计了一种近似理想滤波器来实现深度图的超分辨率。一般来说，基于滤波的 DSR 可以表述如下：

$$D_p = \sum_{q_l \in \Omega_p} (w_{p,q} L_{q_l}) / \sum_{q_l \in \Omega} w_{p,q}$$

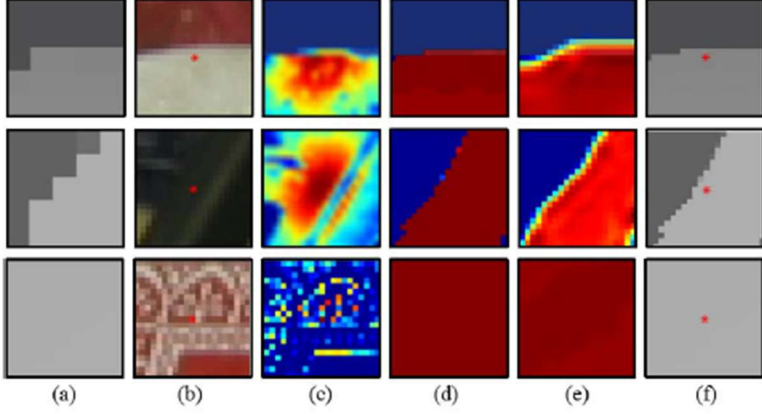


Fig. 3. Filter kernels. The first row presents the database without well-aligned color or depth edges. (a), (b) and (f) are the input depth image, color image, and ground truth depth image, respectively. In (c), the Joint Bilateral Filter (JBF) (with the color as the guidance) can well approximate the *ideal* filter kernel (Refer to Section III-A for details); In (d), kernels computed with the guidance of the ground-truth depth in (f). The color represents the filter weights. Red color corresponds to large weights while blue represents small value. The last two rows present two databases when the color edges are different from the depth edges, it can be seen that the joint bilateral filter kernel will be quite different from (d). While our proposed method can learn a much better approximation of the *ideal* filter kernel as shown in (e).

其中， D 和 L 分别表示上采样的深度图和输入的低分辨率图像， q 表示图像 D 的像素坐标， q_l 描述 q 在图像 L 中的对应坐标。 $w_{p,q}$ 代表以像素 p 为中心的 边缘保留滤波器核 Ω_p 。联合双边滤波器 (joint bilateral filter) 可以被描述为 $w_{p,q} = \exp\left(-\frac{\|p,q\|}{2\sigma_S^2}\right) \exp\left(-\frac{\|G_p - G_q\|}{2\sigma_R^2}\right)$ ，因为当颜色边缘与深度边缘不一致时，联合双边滤波器具有局限性，我们引入了理想的滤波器核。深度卷积神经网络已被证明是非常有效的提取有用的特征，比大多数人工设计的特征具有更好的性能。受此启发，我们打算使用深度 CNN 学习 DSR 的理想滤波器内核。

$$w_p = f_{CNN}(patch_p^H)$$

Color Guidance With High-Resolution Color Image

色彩引导过程主要由五个步骤组成。首先，利用双三次插值方法从 LR 深度图像中获取初始的 HR 深度图像。其次，主要利用 canny 算子检测初始 HR 深度图像的边缘。第三，对 canny 算子检测到的边缘进行膨胀，确定边缘区域。第四，对于初始 HR 深度图像的边缘区域中以像素 p 为中心的邻域，我们在 HR 彩色图像的对应区域中找到一组颜色相似的像素，并确定与像素 p 颜色最相似的像素 q 的位置，最后找到与像素 q 在 LR 深度图像中的位置对应的像素， d_p 为像素 p 的深度值。

让 $patch_p^G$ 表示以像素 p 为中心的高分辨率引导图像 G (例如，彩色图像) 的局部补丁，一个简单的解决方案是将 $patch_p^H$ 和 $patch_p^G$ 都馈送到网络，并且它可以简单地描述为：

$$w_p = f_{CNN}(patch_p^H, patch_p^G)$$

为了减少计算量，我们借鉴了残差网络的思想。通过从 $patch_p^H$ 中减去 d_p 表示的值来修改不确定的插值深度值。理论上， d_p 应该是 $patch_p^H$ 中中心像素 p 的真实深度值，因此引导 patch 可以描述为：

$$Gpatch_p^H = patch_p^H - d_p$$

Deep Coarse-to-Fine Cascade Architecture

如上所述，在颜色引导的情况下，可以通过卷积神经网络来学习近似的理想滤波器核，以重构高分辨率深度

图像。然而，滤波器大小可能与重构的 HR 图像的质量有关系。例如，一些深度值可能只与非常小的邻域中的值相关，而大的核有时可能会影响上采样结果。此外，一般认为卷积层数越多，重建结果越准确。基于这一假设，我们考虑使用深度粗到细网络来增加网络的层数，并改变深度图上采样的内核大小。

我们提出的网络架构主要包括粗阶段和细阶段。在这两个阶段中，卷积神经网络由三个卷积层组成，每个层后面是一个元素激活函数层，它对输入进行线性卷积。每个卷积层具有大小为 $s_1 \times s_2 \times s_3 \times s_4$ 的滤波器组 W 和 s_4 维偏置向量 B 。把一个卷积层及其后面的激活层称为一个块，我们的单个 CNN 有三个块。

$$f_i(x_i) = x_{i+1} = \max(W_i * x_i + B_i, 0)$$

$$\begin{aligned} w_p &= f_{CNN}(Gpatch_p^H) \\ &= f_{CNN}(patch_p^H - d_p) \\ &= f_3\left(f_2\left(f_1(patch_p^H - d_p)\right)\right) \end{aligned}$$

在本文中，我们首先使用 CNN 网络学习一些较大的核来逼近理想的滤波器核。借助于较大的核，获得高分辨率的深度版本。由于边缘区域的一些深度值只与局部小邻域像素有关，我们通过网络学习更小的滤波器核。此外，由于精细 CNN 的输入是分辨率更高的深度图，因此重建的高分辨率深度图像将比输入图像具有更好的质量。

TABLE I
QUANTITATIVE COMPARISON ON THE MIDDLEBURY DATASET 2005 IN MAD WITH THREE UPSAMPLING FACTORS

	Art			Book			Dolls			Laundry			Moebius			Reindeer		
	4×	8×	16×	4×	8×	16×	4×	8×	16×	4×	8×	16×	4×	8×	16×	4×	8×	16×
Bicubic	10.58	19.59	35.15	3.80	8.15	16.41	4.73	9.60	19.49	7.42	14.82	26.77	4.52	9.38	18.51	5.20	9.88	19.13
CLMF0 [54]	7.57	16.72	33.32	3.17	7.25	16.93	3.97	9.65	18.36	6.11	12.57	25.35	4.03	8.40	17.60	4.60	9.71	18.26
CLMF1 [54]	8.12	17.28	33.25	3.27	7.25	16.09	4.04	8.76	18.32	5.50	12.67	25.40	4.13	8.42	17.27	4.65	9.96	18.34
TGV [37]	5.14	10.51	21.37	2.48	4.65	11.20	4.45	11.12	45.54	6.99	16.32	53.61	3.68	6.84	14.09	4.67	11.22	43.48
Guided [18]	9.97	15.53	28.43	3.68	6.52	13.07	4.46	7.63	15.87	6.33	11.90	20.26	4.78	7.88	14.84	5.16	8.11	15.71
JBF [22]	3.36	8.73	21.69	4.05	10.18	19.94	3.98	12.86	29.72	2.39	5.64	13.72	3.19	7.43	15.78	3.89	13.94	27.15
Edge [41]	6.82	13.49	25.90	3.35	8.50	19.32	2.90	6.84	17.97	2.82	5.46	13.57	3.72	7.36	14.05	2.67	6.22	16.80
JGF [53]	3.25	7.39	14.31	2.14	5.41	12.05	3.23	7.29	15.87	2.60	4.54	8.69	3.36	6.45	12.33	2.27	5.17	11.84
AR [52]	4.13	5.58	21.67	1.88	4.16	9.25	4.07	6.62	11.50	3.51	5.19	11.12	2.14	5.57	10.87	3.64	5.76	9.40
Tree [25]	3.96	5.24	9.74	5.77	7.22	11.48	4.60	6.36	13.02	2.27	3.94	8.87	3.52	4.90	8.67	3.97	5.76	2.77
KSVD [34]	3.46	5.18	8.39	2.13	3.97	8.76	4.53	6.18	12.98	2.19	3.89	8.79	2.08	4.86	8.97	2.19	5.76	12.67
CDLLC [55]	2.86	4.59	7.53	1.34	3.67	8.12	4.61	5.94	12.64	2.08	3.77	8.25	1.98	4.59	7.89	2.09	5.39	11.49
JSRD [56]	2.57	4.35	6.79	1.27	3.16	7.93	2.78	5.67	12.19	1.98	2.98	7.98	1.87	4.32	7.64	2.03	4.39	9.83
Xie [26]	2.48	3.31	5.88	1.23	3.09	7.58	2.72	5.59	12.06	1.62	2.86	7.87	1.88	4.29	7.63	1.97	4.31	9.27
PB [57]	3.12	6.18	12.34	1.39	3.34	8.12	3.99	6.22	12.86	2.68	5.62	11.76	1.95	4.12	8.32	6.04	12.17	21.35
SRCNN [7]	7.61	14.54	23.65	2.88	7.98	15.24	3.93	8.34	16.13	6.25	13.63	24.84	3.63	7.28	14.53	3.84	7.98	14.78
ATGV-Net [42]	3.78	3.78	9.68	5.48	7.16	10.32	4.55	6.27	12.64	2.07	3.78	8.69	3.47	4.81	8.56	3.82	5.68	2.63
Song [43]	2.39	3.28	5.82	1.21	2.98	7.48	2.59	5.47	11.78	1.56	2.75	7.64	1.86	4.15	7.52	1.86	3.92	8.67
Wang [35]	7.83	15.21	31.32	3.19	8.52	16.73	4.74	9.53	19.37	6.19	12.86	22.96	3.89	8.23	16.58	3.59	7.23	14.12
MSG-Net [44]	2.31	4.31	8.78	1.21	3.24	7.85	2.39	4.86	9.94	1.68	2.78	7.62	1.79	4.05	7.48	1.73	2.93	7.63
Our_CS	2.28	4.27	8.61	1.35	3.51	8.04	2.01	4.53	10.90	1.55	2.71	7.56	2.25	3.98	7.41	1.59	2.84	7.42
Our_FS	2.23	3.59	7.28	1.19	3.07	7.32	1.98	4.49	9.84	1.39	2.49	7.35	2.18	3.91	7.41	1.51	2.79	6.58

TABLE III

QUANTITATIVE COMPARISON ON THE MIDDLEBURY DATASET 2003 IN *MAD* WITH THREE UPSAMPLING FACTORS

	Tsukuba			Venus			Teddy			Cones		
	2×	4×	8×	2×	4×	8×	2×	4×	8×	2×	4×	8×
Edge [41]	2.35	4.44	6.95	0.44	0.90	2.65	3.12	6.27	13.73	3.26	7.18	14.38
TGV [37]	1.79	3.08	5.31	0.41	0.60	1.76	2.31	3.72	7.51	2.54	4.34	8.17
ScSR [33]	3.27	6.15	9.17	0.71	1.43	3.42	3.76	7.79	15.86	4.43	9.33	17.35
KSVD [34]	2.48	4.30	6.78	0.59	1.22	3.15	2.97	5.17	8.93	3.97	6.45	12.51
SRCNN [7]	2.99	5.52	8.64	0.71	1.30	3.23	3.98	6.92	14.12	4.99	8.64	16.18
CDLLC [55]	2.41	4.15	6.59	0.71	1.18	3.08	2.99	4.72	9.13	3.68	5.79	11.23
Huang <i>et al.</i> [58]	3.53	6.20	9.32	0.67	1.45	3.61	3.88	7.37	15.24	4.52	8.44	15.38
PB [57]	1.57	2.52	3.69	0.39	0.66	1.83	4.13	8.03	17.90	4.35	9.73	17.69
JSRD [56]	1.4	2.37	3.52	0.38	0.59	1.69	1.71	3.13	6.23	1.96	3.23	6.53
Xie [26]	1.27	2.36	3.50	0.37	0.54	1.62	1.61	3.11	6.18	1.72	3.09	6.27
ATGV-Net [42]	1.52	2.41	3.59	0.40	0.63	1.76	5.35	5.37	7.62	4.63	5.74	7.36
Song [43]	1.25	2.23	3.49	0.39	0.53	1.60	1.63	3.10	4.52	1.71	3.05	4.37
Wang [35]	3.12	3.24	5.68	0.68	1.21	2.87	3.92	4.27	5.67	4.83	8.72	9.35
MSG-Net [44]	1.22	2.21	3.44	0.35	0.51	1.58	1.59	3.07	3.69	1.68	2.98	3.73
Our_CS	1.24	2.23	3.46	0.34	0.53	1.62	1.59	3.07	3.67	1.71	2.92	3.71
Our_FS	1.16	2.18	3.42	0.33	0.51	1.56	1.58	2.98	3.58	1.64	2.89	3.70

TABLE IV

QUANTITATIVE COMPARISON ON THE MIDDLEBURY DATASET 2003 IN *PE* WITH THREE UPSAMPLING FACTORS

	Tsukuba			Venus			Teddy			Cones		
	2×	4×	8×	2×	4×	8×	2×	4×	8×	2×	4×	8×
Edge [41]	0.61	0.77	1.32	0.23	0.29	0.56	0.78	1.08	2.13	1.03	1.52	2.98
TGV [37]	0.53	0.71	1.18	0.17	0.24	0.43	0.75	0.83	1.62	0.83	1.13	2.23
ScSR [33]	0.64	0.82	1.62	0.29	0.38	0.64	0.90	1.18	2.31	1.15	1.45	2.84
KSVD [34]	0.51	0.66	1.09	0.23	0.30	0.59	0.70	0.92	2.07	0.91	1.15	2.28
SRCNN [7]	0.64	0.79	1.43	0.28	0.34	0.61	0.88	1.10	2.35	1.12	1.41	2.91
CDLLC [55]	0.48	0.61	0.98	0.21	0.27	0.53	0.67	0.85	1.59	0.85	1.07	2.12
Huang <i>et al.</i> [58]	0.66	0.87	1.73	0.29	0.39	0.69	0.90	1.23	2.68	1.15	1.48	2.88
PB [57]	0.62	0.86	1.71	0.30	0.38	0.62	0.89	1.26	2.73	1.18	1.56	3.11
JSRD [56]	0.47	0.71	1.21	0.18	0.29	0.51	0.64	0.97	1.56	0.81	1.24	2.32
Xie [26]	0.45	0.67	1.09	0.19	0.29	0.49	0.63	0.95	1.51	0.76	1.16	2.14
ATGV-Net [42]	0.46	0.72	0.88	0.23	0.31	0.52	0.69	1.03	1.6	0.83	1.27	2.42
Song [43]	0.43	0.66	0.89	0.17	0.37	0.56	0.68	0.91	1.72	0.75	1.12	2.13
Wang [35]	0.65	0.68	0.83	0.26	0.34	0.69	0.75	1.24	3.01	1.856	1.35	4.86
MSG-Net [44]	0.41	0.62	0.75	0.14	0.34	0.57	0.65	0.82	2.76	0.73	1.06	2.22
Our_CS	0.43	0.65	0.73	0.14	0.26	0.44	0.63	0.80	1.67	0.93	1.08	2.31
Our_FS	0.39	0.61	0.71	0.12	0.25	0.44	0.61	0.79	1.42	0.71	1.05	2.09

TABLE V

QUANTITATIVE COMPARISON ON THE ToFMark DATABASES [37] USING MAD METRIC WITH 4× UPSAMPLING FACTOR

	Tree [25]	PB [57]	SRCNN [7]	GF [18]	JBF [22]	TGV [37]	KSVD [34]	CDLLC [55]	Xie [26]	ATGV-Net [42]	Song [43]	Wang [35]	MSG-Net [44]	Our_CS	Our_FS
Books	5.34	3.91	3.59	3.50	3.38	3.19	3.11	2.86	2.74	2.83	2.64	3.62	2.61	2.54	2.51
Devil	15.07	3.44	2.59	2.47	2.85	2.44	2.32	2.27	2.22	2.16	2.18	2.63	2.11	2.07	2.01
Shark	13.72	5.37	4.48	4.57	3.95	4.07	3.98	3.77	3.46	3.53	3.37	4.67	3.27	3.29	3.26
Average	11.38	4.24	3.56	3.51	3.39	3.23	3.13	2.97	2.79	2.86	2.69	3.64	2.66	2.63	2.60

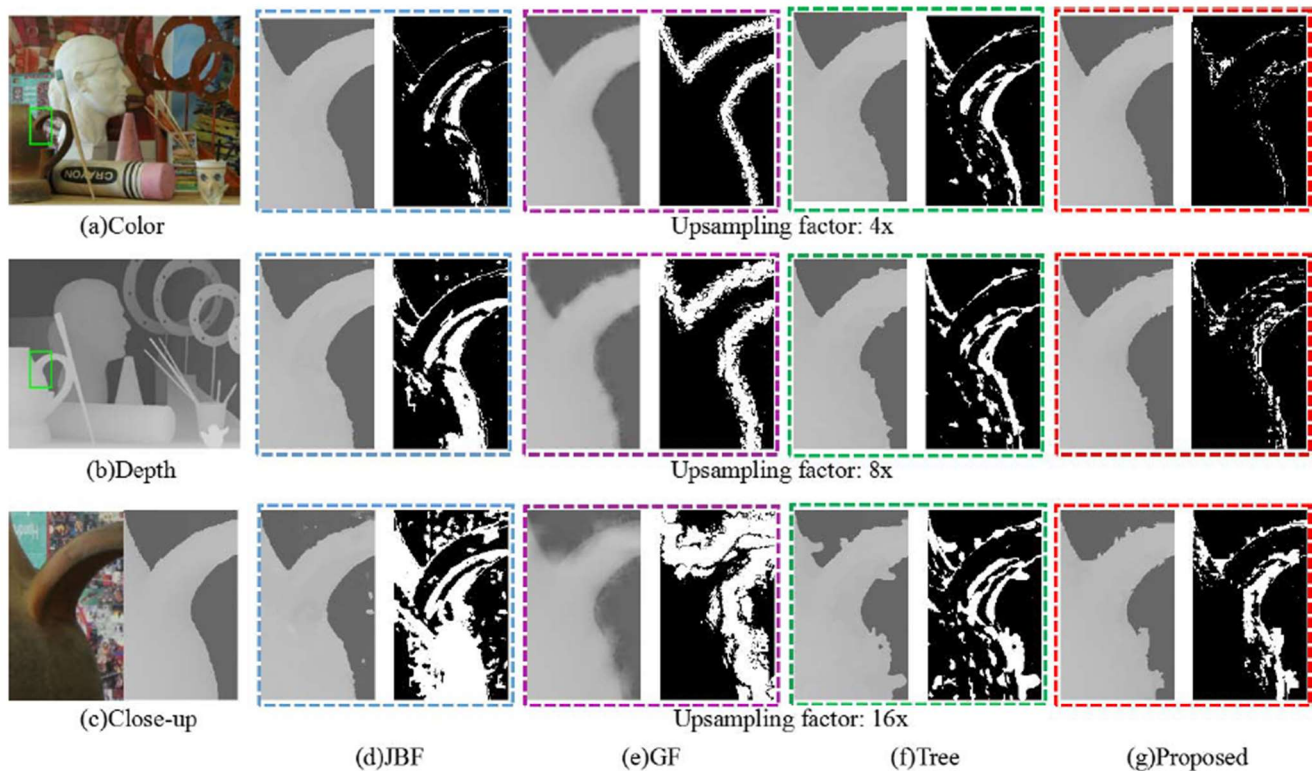


Fig. 8. Edge preserving quality. Visual comparison with popular edge-preserving filtering based upsampling methods [18], [22], [25] were conducted. (a) presents the high-resolution color image and the ground-truth disparity map. (b) and (c) are close-ups from (a). (d)-(g) are the close-ups of the disparity maps upsampled using different methods and the corresponding disparity error maps (obtained with error threshold 1). Note that although edge-preserving filters can all effectively maintain the edges, the accuracy is quite different. As can be seen from the binary error maps, the proposed method achieves the best accuracy around depth discontinuities.