International Journal of Wavelets, Multiresolution
and Information Processing
Vol. 19, No. 2 (2021) 2050072 (17 pages)

DOI: 10.1142/S0219691320500721

Single depth map super-resolution via a deep feedback network

Guoliang Wu*,†,§, Yanjie Wang*,¶ and Shi Li‡,‖

*Changchun Institute of Optics, Fine Mechanics and Physics,
Chinese Academy of Sciences, Changchun 130033, P. R. China

†University of Chinese Academy of Sciences
Beijing 100049, P. R. China

‡CPEEC Electric Power Technical Services Company
Daqing 163711, P. R. China

§wugl@ciomp.ac.cn
¶wangyj@ciomp.ac.cn
‖349115877@qq.com

Received 5 May 2020
Revised 8 September 2020
Accepted 9 October 2020
Published 7 December 2020

Existing depth map-based super-resolution (SR) methods cannot achieve satisfactory results in depth map detail restoration. For example, boundaries of the depth map are always difficult to reconstruct effectively from the low-resolution (LR) guided depth map particularly at big magnification factors. In this paper, we present a novel super-resolution method for single depth map by introducing a deep feedback network (DFN), which can effectively enhance the feature representations at depth boundaries that utilize iterative up-sampling and down-sampling operations, building a deep feedback mechanism by projecting high-resolution (HR) representations to low-resolution spatial domain and then back-projecting to high-resolution spatial domain. The deep feedback (DF) block imitates the process of image degradation and reconstruction iteratively. The rich intermediate high-resolution features effectively tackle the problem of depth boundary ambiguity in depth map super-resolution. Extensive experimental results on the benchmark datasets show that our proposed DFN outperforms the state-of-the-art methods.

Keywords: Depth map; super-resolution; feedback; depth reconstruction.

AMS Subject Classification 2020: 94A08

1. Introduction

Image SR for restoring HR image from LR image, is a long-standing problem in the field of computer vision. Super-resolution techniques as a basic task have been widely used in scene depth recovery,[34] medical imaging,[18] surveillance,[14]

hyper-spectral imaging,[13] biomimetic imaging[24] and so on. With the significant progress of deep learning in recent years, image SR techniques have been greatly improved. Although many approaches are committed to image SR, it is still a difficult challenge as an inherently ill-posed inverse problem.

Depth maps are widely used in various fields, such as autonomous navigation,[22] 3D reconstruction,[9] monitoring[2] and so on. Nevertheless, the problem remains unsolved to acquire high quality and HR depth maps in reality, which requires specialized depth-sensing equipment at great cost in terms of effort and expense. It is difficult or impossible to use depth sensors to obtain HR depth maps directly, thus efficient depth SR techniques are desired to recover HR depth maps from the corresponding data of degraded LR depth map. Recently, several SR methods[40,44] use convolutional neural networks (CNN) to reconstruct HR depth outputs from LR depth inputs. These methods are devoted to resolve the reconstruction of depth map details, e.g. boundaries of depth map, to give a high quality scene depth map. Nevertheless, the recovered depth boundaries generally lose sharpness and are hard to accurately reconstruct from LR depth maps particularly at big magnification factor. In general, these typical methods usually calculate a series of feature maps from LR inputs and then reconstruct HR images through different designed up-sampling modules, such as progressive deconvolutions[44] or combining with texture images to further enhance the feature representations.[40] The core of these methods can be regarded as extracting features step by step with the feedforward information flows from LR images to the final SR images.

In cognitive theory, the transmission channels of the response signal from the higher-order region to the lower-order region are composed of the feedback connection of the visual region of the cerebral cortex[10,25] which are indispensable in expression and regulation of human. If we consider the LR-to-HR mapping as a feedforward mechanism, then the opposite HR-to-LR mapping can be considered as a feedback process where the rich intermediate HR features can be effectively extracted to tackle the puzzle of limited spatial resolution and depth boundary ambiguity in LR images. Different from the above simple feedforward methods, some works[4,26,35,45] have proved the effect of feedback mechanism in some deep learning-based tasks. Therefore, inspired by the feedback mechanism, we present a novel SR method for single depth map by introducing a deep feedback network, which can effectively enhance the feature representations at depth boundaries that utilize iterative up-sampling and down-sampling operations, building a deep feedback mechanism by projecting HR representations to LR spatial domain and then back-projecting to HR spatial domain. The designed deep feedback blocks transmit the extracted HR features back to the previous LR feature domain to refine the LR feature map. The key point of feedback mechanism is that the coarse HR information can provide crucial information to guide the reconstruction of high quality depth maps. To maximize the use of hierarchical features, we utilize dense connections to connect all the deep feedback blocks to form a dense DFN, which can further improve the performance.

Our main contributions are summarized in the following:

(1) An end-to-end learning framework for high-quality and HR depth map SR from LR depth input, which takes full advantage of all hierarchical features of the original LR depth map.
(2) A DF designed based on the feedback mechanism, which can not only efficiently handle feedforward and backward information flows, but also enrich the feature representations to promote the recovery of high quality depth map.
(3) Extensive experimental results on the benchmark datasets show that our proposed DFN outperforms the state-of-the-art methods. Note that, different from other methods that use texture image as auxiliary information to promote the recovery of the sharp depth boundaries, we only focus on singe depth map SR, but achieve superior performance surprisingly.

2. Related Work

The significant progress of deep learning has greatly facilitated the progress of image SR. Dong *et al.*[5] successfully solved image SR tasks by CNN, more and more researches adopt deep learning to solve the problem of image SR. We present a simple overview of image SR based on deep learning in Sec. 2.1. The feedback mechanism has been proven to have excellent error feedback ability in existing deep learning-based methods and has practically promoted the SR field. We present a briefly overview in Sec. 2.2.

2.1. *Deep learning-based super-resolution*

Dong *et al.*[5] first built a deep learning-based SR method, i.e. SRCNN, which is only a three-layer CNN structure and outperforms traditional methods. SRCNN uses interpolation to up-sample LR, then uses the middle result as the input of CNN. Based on SRCNN, SFRCNN[6] and ESPCN[36] up-sampled LR at the end of the networks after extracting a series of feature maps, which can effectively increase the resolution and reduce middle operations. Obviously, deeper networks generally outperform shallower networks. VDSR[23] and DRCN[21] significant improved the accuracy by increasing the depth of the network. EDSR[29] and SRResNet[29] took advantage of residual learning and used the efficient sub-pixel convolution layer to further extract more global and local features. With the help of dense connections,[17] DSC[37] and RDN[47] build deeper networks which had more parameters at the same time. Some works[1,42] have been carried out to consider the tradeoff between performance and network parameters.

Encouraged by the advance of single image SR, some researches apply deep learning to the special tasks of depth map-based SR. Riegler *et al.*[34] presented an end-to-end framework for reconstructing HR depth output from corresponding LR depth input in two paths. Ferstl *et al.*[8] presented a variational up-sampling network which used the representations of example-based edge. Gu *et al.*[11] presented

an analysis representation model that can enhance the depth map and solve the limitation of the complex interdependence between depth map and color image. Hui *et al.*[19] presented a method which introduced gradual upsampling and multi-scale color guidance. They further explored the relationship between depth structure and color texture. Ye *et al.*[44] presented a depth-based SR network by learning a binary feature of depth boundaries from HR color image and LR depth map. Wen *et al.*[40] also used the color guidance and proposed a cascade network by color guidance to eliminate artifacts from texture copying.

2.2. *Feedback network*

Recently, feedback mechanism has been applied to various computing tasks.[4,26,35,38] The feedback mechanism makes the deep learning network to learn more hierarchical information. Carreira *et al.*[4] presented an iterative error feedback model by estimating and correcting the current results iteratively. PreNet[31] proposed an unsupervised recurrent network by predicting the future frames and importing the estimation back to the network.

The feedback mechanism has demonstrated exciting capability in the area of image SR. Han *et al.*[12] utilized the delayed feedback mechanism in a dual-state recurrent neural network by transmitting the information flow between two recurrent states.

In general, the above-introduced SR methods utilized CNN to learn nonlinear LR-to-HR mapping, but ignored the interrelation between HR image and LR image. Compared to the above methods, our DFN models the process of image degradation and reconstruction and utilizes iterative up-sampling and down-sampling, introducing a deep feedback mechanism by projecting HR representations to LR spatial domain.

3. Proposed Method

3.1. *Network structure*

In this section, the pipeline of our network architecture is illustrated as follows: shallow feature extraction block, deep feedback block, and reconstruction block, as shown in Fig. 1. To express the implementation details, we use colored rectangles to represent the different execution processes in each block. We denote the LR depth map input and the SR depth map output from our DFN as I_{LR} and I_{SR}, respectively. We first send a LR input I_{LR} to the shallow feature extraction block, from which we can acquire the initial features I_{SF} by two continuous convolutional layers:

$$I_{\mathrm{SF}} = F_{\mathrm{SFB}}(I_{\mathrm{LR}}), \tag{3.1}$$

where F_{SFB} denotes the feature extraction operation in the shallow feature extraction block. I_{SF} is then used for the following stacked deep feedback blocks. So we

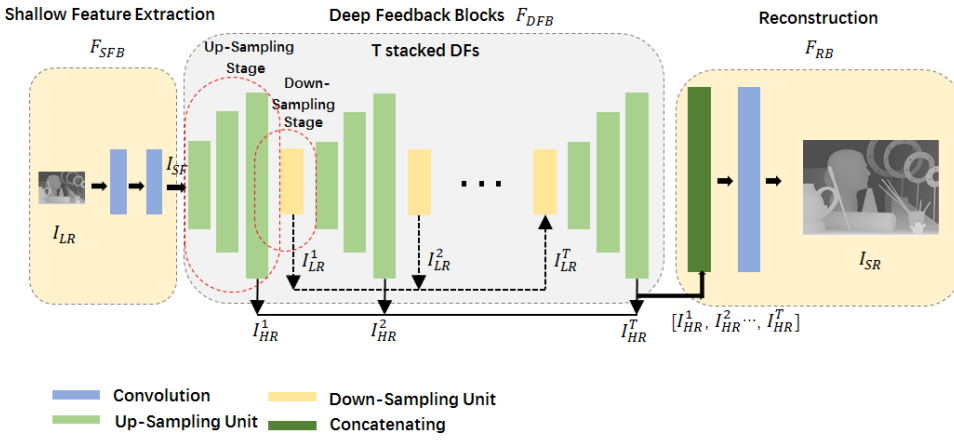


Fig. 1. Network architecture of the proposed DFN.

can further have

$$[I_{\mathrm{HR}}^1, I_{\mathrm{HR}}^2, \ldots, I_{\mathrm{HR}}^T] = F_{\mathrm{DFB}}(I_{\mathrm{SF}}), \tag{3.2}$$

where F_{DFB} denotes our proposed deep feedback mechanism, which contains T up-sampling stages and $T-1$ down-sampling stages. Each I_{HR} is produced in each up-sampling stage, so we get T high resolution feature maps totally. Note that we design a step-by-step up-sampling strategy in our up-sampling stage, which contains multiple progressive up-sampling units in terms of large upscaling factors, while only one step to downsample the feature map to the low resolution domain. We will give more details about the up-sampling unit and down sampling unit in our DF structure in the following section. Finally, the target SR image I_{SR} is reconstructed from the multiple HR feature maps $[I_{\mathrm{HR}}^1, I_{\mathrm{HR}}^2, \ldots, I_{\mathrm{HR}}^T]$:

$$I_{\mathrm{SR}} = F_{\mathrm{RB}}([I_{\mathrm{HR}}^1, I_{\mathrm{HR}}^2, \ldots, I_{\mathrm{HR}}^T]), \tag{3.3}$$

where $[\cdot]$ denotes concatenation operation. We concatenate each output of each up-sampling stage and then get target channels by a convolutional layer. F_{RB} denotes the operations of reconstruction block.

3.2. *Deep feedback block*

More details about our proposed DF are provided in this section. As shown in Fig. 1, DF block contains T up-sampling stage and $T-1$ down-sampling stage, and each sampling operation is implemented by a up-sampling unit or a down-sampling unit. The DF is designed to exploit iterative up-sampling and down-sampling operations, introducing a deep feedback mechanism by projecting HR representations to LR spatial domain and then back-projecting to HR spatial domain. The rich intermediate HR features effectively solve the problem of depth boundary ambiguity in depth map SR.

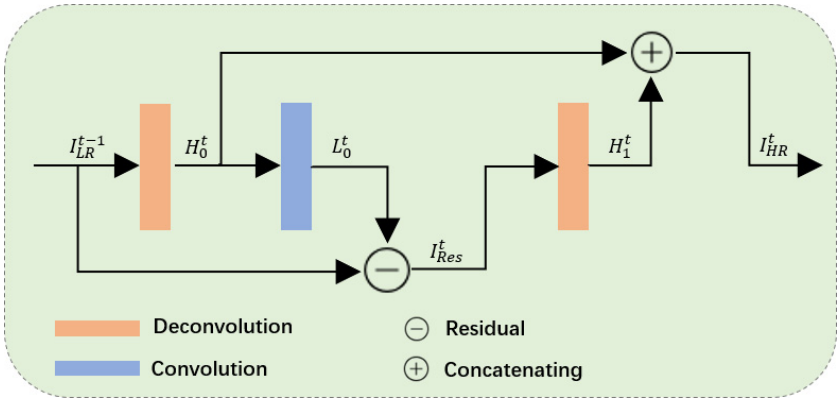


Fig. 2. The up-sampling unit.

Up-sampling unit

We denote I_{HR}^t and I_{LR}^t as the feature output of up-sampling unit and down-sampling unit, respectively. The up-sampling unit is designed in Fig. 2. We treat the output I_{LR}^{t-1} of $(t-1)$th down-sampling unit as input at the tth iteration, then project it to HR features H_0^t. H_0^t can be obtained by

$$H_0^t = Up_0^t(I_{\mathrm{LR}}^{t-1}), \tag{3.4}$$

where Up_0^t is the first deconvolution operation. Then we would like to map H_t^0 back to LR features L_0^t, which can be obtained by

$$L_0^t = \mathrm{Down}_0^t(H_0^t), \tag{3.5}$$

where Down_0^t is a convolution operation. The residual I_{Res}^t is calculated between I_{LR}^{t-1} and L_0^t:

$$I_{\mathrm{Res}}^t = I_{\mathrm{LR}}^{t-1} - L_0^t. \tag{3.6}$$

The residual I_{Res}^t is mapped to HR features H_1^t by the second deconvolution operation:

$$H_1^t = Up_1^t(I_{\mathrm{Res}}^t). \tag{3.7}$$

The final output I_{SR}^t of up-sampling unit can be obtained by

$$I_{\mathrm{HR}}^t = H_0^t + H_1^t. \tag{3.8}$$

At the up-sampling step, our progressive up-sampling strategy uses a pyramidal structure which gradually amplifies the resolution of feature maps to the desired size. For example, when dealing with $8x$ up-sampling case, we need three up-sampling units, each containing a up-sampling operation with a $2x$ upscale factor. This strategy can avoid using large-sized deconvolution filters and increase the effectiveness of training.

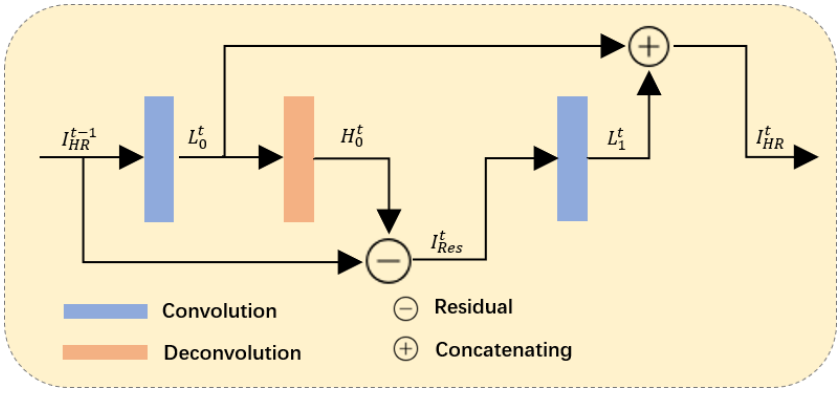


Fig. 3. The down-sampling unit.

Down-sampling unit

The down-sampling unit can be considered as an inverse process of up-sampling unit, which is illustrated in Fig. 3, and we present the corresponding formulation flows as follows:

$$L_0^t = \mathrm{Down}_0^t(I_{\mathrm{HR}}^{t-1}). \tag{3.9}$$

Similarly, after the first convolution operation, we would like to map L_0^t to SR features H_0^t, which can be obtained by

$$H_0^t = Up_0^t(L_0^t). \tag{3.10}$$

The residual I_{Res}^t is calculated between I_{HR}^t and H_0^t:

$$I_{\mathrm{Res}}^t = I_{\mathrm{HR}}^{t-1} - H_0^t. \tag{3.11}$$

The residual I_{Res}^t is mapped to LR features L_1^t by the second convolution operation:

$$L_1^t = \mathrm{Down}_1^t(I_{\mathrm{Res}}^t). \tag{3.12}$$

The final output I_{LR}^t of down-sampling unit can be obtained by

$$I_{\mathrm{LR}}^t = L_0^t + L_1^t. \tag{3.13}$$

In contrast to up-sampling unit, down-sampling unit aims to project the input HR feature map I_{HR} to the LR I_{LR} domain, as illustrated in Fig. 3.

Dense Connections

According to the pioneering work, dense connection is widely introduced in various area of computer vision tasks,[15–17,20,46,47] which relieves the vanishing-gradient puzzle and takes advantage of all hierarchical features. Therefore, as shown in Fig. 1, the concatenation of all previous output units can be used as the input of each

up-sampling unit and down-sampling unit. $\bar{I}^t_{\mathrm{LR}}$ and $\bar{I}^t_{\mathrm{HR}}$, respectively, denote the input of up-sampling unit and down-sampling unit. $\bar{I}^t_{\mathrm{LR}}$ and $\bar{I}^t_{\mathrm{HR}}$ can be obtained by

$$\bar{I}^t_{\mathrm{LR}} = [\bar{I}^1_{\mathrm{LR}}, \bar{I}^2_{\mathrm{LR}}, \ldots, \bar{I}^{t-1}_{\mathrm{LR}}], \tag{3.14}$$

$$\bar{I}^t_{\mathrm{HR}} = [\bar{I}^1_{\mathrm{HR}}, \bar{I}^2_{\mathrm{HR}}, \ldots, \bar{I}^{t-1}_{\mathrm{HR}}]. \tag{3.15}$$

4. Results

4.1. *Settings*

Dataset and Metrics. We use the most popular Middlebury dataset[a] as our training and test data according to the official splitting (38 images for training and 6 for validation), and use MPI Sintel dataset[b] to test the generalization of our proposed method. To generate LR depth inputs, HR depth maps are degraded on target size using Bicubic. We augmented the train dataset by flipping horizontally and randomly extracted more than 10,000 depth patches of a fixed size 15×15 from down-sampled depth map. According to 2, 4, 8 and 16 up-sampling factors, the squared sizes of corresponding HR depth patches are 30, 60, 120 and 240, respectively. We also evaluate SR results under the mean absolute difference (MAD) metric on six standard test depth maps as other methods, i.e. *Art*, *Books*, *Moebius*, *Dolls*, *Laundry*, *Reindeer*.

Implementation Details. For each magnification factor, we used the kernel size of 6×6 with a stride size of 2 at each up-sampling unit. At down-sampling stage, we used the kernel size of 6×6, 8×8, 12×12, and 20×20 for $2\times, 4\times, 8\times$ and $16\times$ cases, respectively. Each up-sampling unit or down-sampling unit outputs 64 channels and each convolutional or deconvolutional layer is followed by PReLU. We trained our network by ADAM optimizer with L1 loss and with a batch-size of 16. The initial learning rate is 0.0001 and multiplied by 0.1 per 100 epochs. PyTorch framework was implemented with a NVIDIA 1080Ti GPU.

4.2. *Performance comparisons*

With the aim of verifying the capability of our proposed DFN, we compared our method with other 16 state-of-the-art depth SR methods on $2\times, 4\times, 8\times$ and $16\times$ upscaling factors, including CLMF1,[32] JGF,[30] Edge,[33] IMLS,[3] TGV,[7] AR,[43] FGI,[28] EG,[41] DJF,[27] ATGVNet,[34] DSP,[39] MSG,[19] DGDIE,[11] CCFN,[40] DEIN.[44] Quantitative depth SR results are illustrated in Table 1, which test on Middlebury datasets at four subsampling rate about six testing depth maps. Note that the methods from CLMF1 to DJF are traditional filter-based or optimization-based techniques, which obtain lower MAD values when compared to deep learning-based methods. Compared with the methods based on CNN, the DFN proposed by us almost gets the

[a]Middlebury datasets, http://vision.middlebury.edu/.
[b]MPI Sintel datasets, http://sintel.is.tue.mpg.de/.

Table 1. Quantitative depth up-sampling results on Middlebury datasets at four subsampling rates.

	Art				Book				Dolls			
	2×	4×	8×	16×	2×	4×	8×	16×	2×	4×	8×	16×
Bicubic	0.48	0.97	1.85	3.59	0.13	0.29	0.59	1.15	0.20	0.36	0.66	1.18
CLMF1	0.44	0.76	1.44	2.87	0.14	0.28	0.51	1.02	0.23	0.34	0.60	1.01
JGF	0.29	0.47	0.78	1.54	0.15	0.24	0.43	0.81	0.19	0.33	0.59	1.06
Edge	0.41	0.65	1.03	2.11	0.17	0.30	0.56	1.03	0.16	0.31	0.56	1.05
IMLS	0.27	0.68	1.04	2.20	0.16	0.26	0.48	1.16	0.24	0.36	0.61	0.98
TGV	0.45	0.65	1.17	2.30	0.18	0.27	0.42	0.82	0.21	0.33	0.70	2.20
AR	0.18	0.49	0.64	2.01	0.12	0.22	0.37	0.77	0.21	0.34	0.50	0.82
FGI	0.70	1.29	2.41	4.51	0.43	0.74	1.16	1.91	0.54	0.93	1.44	2.12
EG	—	0.64	—	—	—	0.28	—	—	—	0.33	—	—
DJF	0.12	0.40	1.07	2.78	**0.05**	0.16	0.45	1.00	**0.06**	0.20	0.49	0.99
ATGVNet	—	0.65	0.81	1.42	—	0.43	0.51	0.79	—	0.41	0.52	0.56
DSP	—	0.73	1.56	3.03	—	0.28	0.61	1.31	—	0.32	0.65	1.45
MSG	—	0.46	0.76	1.53	—	0.15	0.41	0.76	—	0.25	0.51	0.87
DGDIE	0.20	0.48	1.20	2.44	0.14	0.30	0.58	1.02	0.16	0.34	0.63	0.93
CCFN	—	0.43	0.72	1.50	—	0.17	0.36	0.69	—	0.25	0.46	0.75
DEIN	0.23	0.40	0.64	**1.34**	0.12	0.22	0.37	0.78	0.12	0.22	0.38	**0.73**
DFN	**0.10**	**0.24**	**0.59**	1.55	0.08	**0.14**	**0.28**	**0.57**	0.09	**0.18**	**0.36**	0.74

Table 1. (*Continued*)

	Laundry				Moebius				Reindeer			
	2×	4×	8×	16×	2×	4×	8×	16×	2×	4×	8×	16×
Bicubic	0.28	0.54	1.04	1.95	0.13	0.30	0.59	1.13	0.30	0.55	0.99	1.88
CLMF1	0.30	0.50	0.80	1.67	0.15	0.29	0.51	0.97	0.32	0.51	0.84	1.55
JGF	0.21	0.36	0.64	1.20	0.15	0.25	0.46	0.80	0.23	0.38	0.64	1.09
Edge	0.17	0.32	0.54	1.14	0.18	0.29	0.51	1.10	0.20	0.37	0.63	1.28
IMLS	0.23	0.39	0.81	1.53	0.15	0.25	0.49	0.93	0.32	0.64	0.74	1.43
TGV	0.31	0.55	1.22	3.37	0.18	0.29	0.49	0.90	0.32	0.49	1.03	3.05
AR	0.20	0.34	0.53	1.12	0.10	0.20	0.40	0.79	0.22	0.40	0.58	1.00
FGI	0.51	0.91	1.59	2.68	0.42	0.72	1.13	1.81	0.50	0.87	1.58	2.72
EG	—	0.37	—	—	—	0.29	—	—	—	0.40	—	—
DJF	0.07	0.28	0.71	1.67	**0.06**	0.18	0.46	1.02	**0.07**	0.23	0.60	1.36
ATGVNet	—	0.37	0.89	0.94	—	0.38	0.45	0.80	—	0.41	0.58	1.01
DSP	—	0.45	0.98	2.01	—	0.31	0.59	1.26	—	0.42	0.84	1.73
MSG	—	0.30	0.46	1.12	—	0.21	0.43	0.76	—	0.31	0.52	0.99
DGDIE	0.15	0.35	0.86	1.56	0.14	0.28	0.58	0.98	0.16	0.35	0.73	1.29
CCFN	—	0.24	0.41	**0.71**	—	0.23	0.39	0.73	—	0.29	0.46	0.95
DEIN	0.13	0.23	**0.36**	0.81	0.11	0.20	0.35	0.73	0.15	0.26	0.40	**0.80**
DFN	**0.07**	**0.16**	0.39	1.16	0.07	**0.14**	**0.28**	**0.66**	0.09	**0.17**	**0.34**	0.94

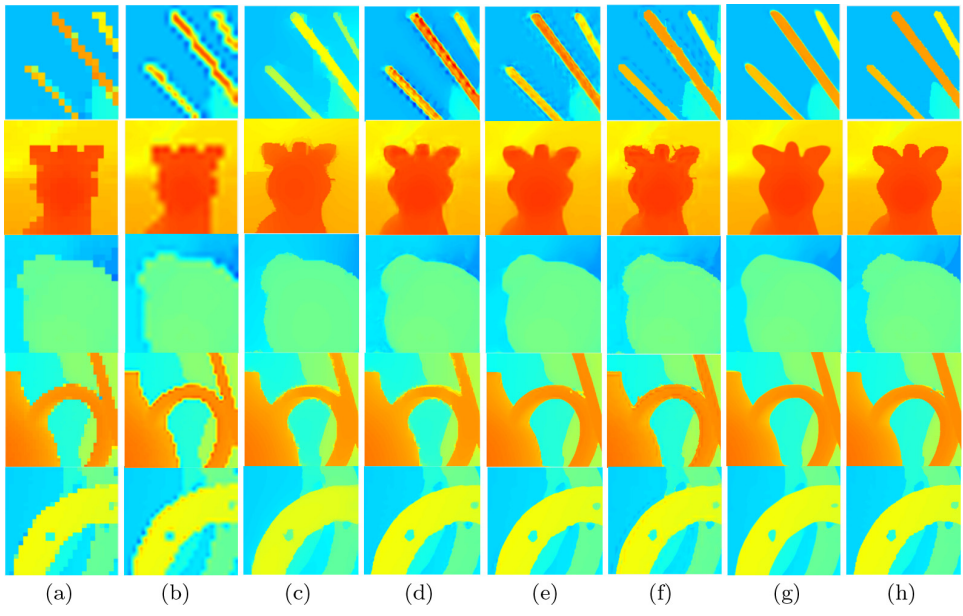


Fig. 4. Qualitative comparision of our method with works on 8×. (a) LR; (b) Bicubic; (c) FGI; (d) DJF; (e) DGDIE; (f) DEIN; (g) DFN; (h) Ground truth.

best objective result. Figure 4 ulteriorly illustrates the visual performance of our method.

4.3. *Ablation investigation*

Ablation study on the key modules of our framework is validated in this section.

Analysis of feedback blocks (DFs). With the aim of verifying the capability of DF, we explore the impact of the number of deep feedback blocks (a group of up-sampling stage and down-sampling stage is treated as one DF block), which is denoted as G. We construct multiple DF blocks under different numbers of G ($G = 1$, $G = 2$, $G = 4$, $G = 6$). We implemented these models under the up-scaling factor of 8×. Quantitative compared results are shown in Table 2. The performance improves as G gets larger. However, the performance of the model approaches saturation when G approaches 6, i.e. $G = 4$ and $G = 6$ have little difference in the results. Therefore, the case of $G = 6$ is our final choice in our DFN.

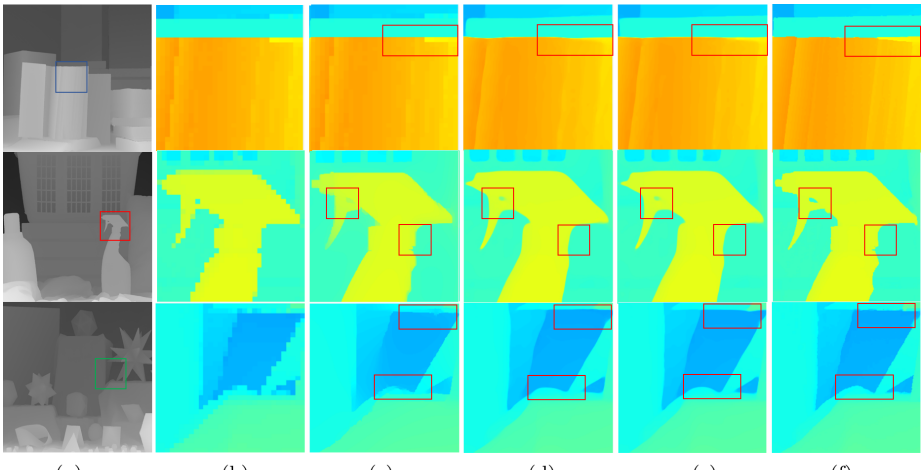
Analysis of different input configurations. In order to further demonstrate the effects of the proposed feedback mechanism, the first up-sampling unit is replaced with bicubic interpolation. We regard the interpolation as the up-sampling operation for generating the middle results which are then fed to our DFN as input. The quantitative results are compared in Table 3. The case without the first up-sampling

Table 2. Ablation investigation on different numbers of feedback blocks on 8×.

	Art	Book	Dolls	Laundry	Moebius	Reindeer
$G = 1$	0.75	0.37	0.39	0.48	0.34	0.47
$G = 2$	0.62	0.29	0.36	0.41	0.29	0.38
$G = 4$	**0.58**	**0.28**	**0.36**	**0.40**	**0.28**	**0.37**
$G = 6$	0.59	**0.28**	**0.36**	**0.40**	**0.28**	**0.37**

Table 3. Ablation investigation on different input configurations.

	Art				Book				Dolls			
	2×	4×	8×	16×	2×	4×	8×	16×	2×	4×	8×	16×
DFN-I	0.15	0.33	0.72	1.95	0.08	0.16	0.31	0.77	0.11	0.21	0.39	0.81
DFN	**0.096**	**0.24**	**0.59**	**1.55**	**0.08**	**0.14**	**0.28**	**0.57**	**0.09**	**0.18**	**0.36**	**0.74**
	Laundry				Moebius				Reindeer			
	2×	4×	8×	16×	2×	4×	8×	16×	2×	4×	8×	16×
DFN-I	0.09	0.21	0.47	1.30	0.08	0.17	0.33	0.76	0.11	0.26	0.45	1.18
DFN	**0.07**	**0.16**	**0.39**	**1.16**	**0.07**	**0.14**	**0.28**	**0.66**	**0.09**	**0.17**	**0.34**	**0.94**



(a) (b) (c) (d) (e) (f)

Fig. 5. (a) The original *book*, *laundry and moebius* examples; (b) LR; (c) FGI; (d) DFN-I; (e) DFN; (f) Ground truth.

unit (DFN-I) achieves higher MAD values than the DFN. As shown in Fig. 5, the DFN also shows the better visual 8× depth SR results.

4.4. *Generalization investigation*

Considering the generalization of our DFN model, another four depth maps picked from MPI depth dataset are tested. Insisting on a point that the MPI dataset

Table 4. Comparison with the-state-of-the-arts.

	Ambush 2–15				Ambush 4–12				Ambush 5–41				Temple 3–23			
	2×	4×	8×	16×	2×	4×	8×	16×	2×	4×	8×	16×	2×	4×	8×	16×
EG	—	0.28	—	—	—	0.76	—	—	—	0.88	—	—	—	0.54	—	—
DJF	0.06	0.20	0.48	0.96	0.21	0.54	1.14	2.49	0.28	0.72	1.42	2.67	0.15	0.40	0.79	1.76
DGDIE	0.23	0.21	0.65	1.24	0.23	0.57	1.26	2.23	0.23	0.73	1.79	3.10	0.23	0.40	1.01	1.90
DEIN	0.09	0.15	**0.28**	**0.54**	0.25	0.50	0.82	1.76	0.24	**0.41**	**0.69**	**1.32**	0.17	0.30	0.51	1.12
DFN	**0.06**	**0.14**	0.30	0.75	**0.12**	**0.27**	**0.58**	**1.29**	**0.19**	0.47	1.06	2.19	**0.05**	**0.12**	**0.27**	**0.50**

(a) (b) (c) (d)

Fig. 6. Qualitative comparison of our method with works on 8×. (a) The Ground truth of *Temple*3-23 *and Alley*1-48 examples; (b) DJF; (c) DEIN; (d) ours.

is obviously different from our training dataset Middlebury. Table 4 presents the performance of generalization investigation compared with EG,[41] DJF,[27] DGDIE and DEIN[44] and demonstrates the generalization ability of our model. Figure 6 further confirms that the method is also excellent in visual performance on MPI dataset, and we box out the salient areas.

5. Conclusions

In this paper, we present a novel SR method for single depth map by introducing a deep feedback network unlike the previous networks which reconstruct the final SR image from the LR depth map in a feedforward manner. Our proposed network aims to directly project HR representations to LR spatial utilizing iterative up-sampling and down-sampling. The deep feedback block iteratively imitates the process of image degradation and reconstruction where the rich hierarchical HR features effectively solve the trouble of ambiguity in depth map SR. Extensive experimental results have shown that our method almost gets the best objective result under the MAD metric and shows excellent visual quality compared with the state-of-the-art methods. Our framework also achieves good performance in terms of ablation and generalization.

Acknowledgments

We earnestly appreciate the work of Editors/Reviewers, and acknowledge their excellent and professional comments and suggestions.

References

1. N. Ahn, B. Kang and K. Sohn, Fast, accurate, and lightweight super-resolution with cascading residual network, *Computer Vision-ECCV 2018-15th European Conf. Proc. Part X*, Munich, Germany, September 8–14, 2018, pp. 256–272.

2. G. Borghi, Combining deep and depth: Deep learning and face depth maps for driver attention monitoring, *CoRR*, arXiv:abs/1812.05831.
3. N. K. Bose and N. A. Ahuja, Superresolution and noise filtering using moving least squares, *IEEE Trans. Image Process.* **15**(8) (2006) 2239–2248.
4. J. Carreira, P. Agrawal, K. Fragkiadaki and J. Malik, Human pose estimation with iterative error feedback, *2016 IEEE Conf. Computer Vision and Pattern Recognition*, 2016, Las Vegas, NV, USA, June 27–30 2016, pp. 4733–4742.
5. C. Dong, C. C. Loy, K. He and X. Tang, Image super-resolution using deep convolutional networks, *IEEE Trans. Pattern Anal. Mach. Intell.* **38**(2) (2016) 295–307.
6. C. Dong, C. C. Loy and X. Tang, Accelerating the super-resolution convolutional neural network, *14th European Conf. Computer Vision-ECCV Proc., Part II*, Amsterdam, The Netherlands, October 11–14 2016, pp. 391–407.
7. D. Ferstl, C. Reinbacher, R. Ranftl, M. Rüther and H. Bischof, Image guided depth upsampling using anisotropic total generalized variation, *Proc. ICCV*, Sydney, Australia, December 1–8, 2013.
8. D. Ferstl, M. Rüther and H. Bischof, Variational depth superresolution using example-based edge representations, *2015 IEEE Int. Conf. Computer Vision*, Santiago, Chile, December 7–13 2015, pp. 513–521.
9. A. Geiger, J. Ziegler and C. Stiller, Stereoscan: Dense 3d reconstruction in real-time, *IEEE Intelligent Vehicles Symp. (IV)* , Baden-Baden, Germany, June 5–9 2011, pp. 963–968.
10. S. Grossberg and J. A. Marshall, Stereo boundary fusion by cortical complex cells: A system of maps, filters, and feedback networks for multiplexing distributed data, *Neural Network.* **2**(1) (1989) 29–51.
11. S. Gu, W. Zuo, S. Guo, Y. Chen, C. Chen and L. Zhang, Learning dynamic guidance for depth image enhancement, *2017 IEEE Conf. Computer Vision and Pattern Recognition*, Honolulu, HI, USA, July 21–26 2017, pp. 712–721.
12. W. Han, S. Chang, D. Liu, M. Yu, M. Witbrock and T. S. Huang, Image super-resolution via dual-state recurrent networks, *2018 IEEE Conf. Computer Vision and Pattern Recognition*, Salt Lake City, UT, USA, June 18–22 2018, pp. 1654–1663.
13. X. Han, J. Wang, B. Shi, Y. Zheng and Y. Chen, Hyper-spectral image super-resolution using non-negative spectral representation with data-guided sparsity, *19th IEEE Int. Symp. Multimedia*, Taichung, Taiwan, December 11–13 2017, pp. 500–506.
14. Q. He, R. R. Schultz and C. H. H. Chu, Efficient super-resolution image reconstruction applied to surveillance video captured by small unmanned aircraft systems, *Proc. Spie* **6968** (2008) 696814–696814–11.
15. G. Huang, D. Chen, T. Li, F. Wu, L. Van Der Maaten and K. Q. Weinberger, Multi-scale dense convolutional networks for efficient prediction, arXiv:1703.09844.
16. G. Huang, S. Liu, L. Van der Maaten and K. Q. Weinberger, Condensenet: An efficient densenet using learned group convolutions, *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, Salt Lake City, UT, USA, June 18–22, 2018, pp. 2752–2761.
17. G. Huang, Z. Liu, L. van der Maaten and K. Q. Weinberger, Densely connected convolutional networks, *2017 IEEE Conf. Computer Vision and Pattern Recognition*, Honolulu, HI, USA, July 21–26 2017, pp. 2261–2269.
18. L. Huang, Y. Labyed, K. Hanson, D. Sandoval, J. Pohl and M. Williamson, Detecting breast microcalci fications using super-resolution ultrasound imaging: A clinical study, *Medical Imaging 2013: Ultrasonic Imaging, Tomography, and Therapy*, International Society for Optics and Photonics, 2013.

19. T. Hui, C. C. Loy and X. Tang, Depth map super-resolution by deep multi-scale guidance, *Computer Vision-ECCV 2016-14th European Conf., Proc. Part III*, Amsterdam, The Netherlands, October 11–14 2016, pp. 353–369.
20. S. Jégou, M. Drozdzal, D. Vazquez, A. Romero and Y. Bengio, The one hundred layers tiramisu: Fully convolutional densenets for semantic segmentation, *Proc. IEEE Conf. Computer Vision and Pattern Recognition Workshops*, 2017, pp. 11–19.
21. Jiwon, J. K. Lee and K. M. Lee, Deeply-recursive convolutional network for image super-resolution, *2016 IEEE Conf. Computer Vision and Pattern Recognition*, Las Vegas, NV, USA, June 27–30 2016, pp. 1637–1645.
22. C. Kerl, J. Sturm and D. Cremers, Dense visual SLAM for RGB-D cameras, *2013 IEEE/RSJ Int. Conf. Intelligent Robots and Systems*, Tokyo, Japan, November 3–7 2013, pp. 2100–2106.
23. J. Kim, J. K. Lee and K. M. Lee, Accurate image super-resolution using very deep convolutional networks, *2016 IEEE Conf. Computer Vision and Pattern Recognition*, Las Vegas, NV, USA, June 27–30 2016, pp. 1646–1654.
24. K. Kim, K.-W. Jang, J.-K. Ryu and K.-H. Jeong, Deep subwavelength nanometric image reconstruction using fourier domain optical normalization, *Light Sci. Appl.* **9** (2020) 28.
25. S. Kriz and M. Hegarty, Top-down and bottom-up influences on learning from animations, *Int. J. Man-Mach. Stud.* **65**(11) (2007) 911–930.
26. K. Li, B. Hariharan and J. Malik, Iterative instance segmentation, *2016 IEEE Conf. Computer Vision and Pattern Recognition*, Las Vegas, NV, USA, June 27–30 2016, pp. 3659–3667.
27. Y. Li, J. Huang, N. Ahuja and M. Yang, Joint image filtering with deep convolutional networks, *CoRR* arXiv:abs/1710.04200.
28. Y. Li, D. Min, M. N. Do and J. Lu, Fast guided global interpolation for depth and motion, *Computer Vision-ECCV 2016-14th European Conf. Proc. Part III*, Amsterdam, The Netherlands, October 11–14 2016, pp. 717–733.
29. B. Lim, S. Son, H. Kim, S. Nah and K. M. Lee, Enhanced deep residual networks for single image super-resolution, *2017 IEEE Conf. Computer Vision and Pattern Recognition Workshops*, Honolulu, HI, USA, July 21–26 2017, pp. 1132–1140.
30. M. Liu, O. Tuzel and Y. Taguchi, Joint geodesic upsampling of depth images, *2013 IEEE Conf. Computer Vision and Pattern Recognition*, Portland, OR, USA, June 23–28 2013, pp. 169–176.
31. W. Lotter, G. Kreiman and D. Cox, Deep Predictive Coding Networks for Video Prediction and Unsupervised Learning, arXiv:abs/1605.08104.
32. J. Lu, K. Shi, D. Min, L. Lin and M. N. Do, Cross-based local multipoint filtering, *2012 IEEE Conf. Computer Vision and Pattern Recognition*, Providence, RI, USA, June 16–21 2012, pp. 430–437.
33. J. Park, H. Kim, Y. Tai, M. S. Brown and I. Kweon, High quality depth map upsampling for 3d-tof cameras, *IEEE Int. Conf. Computer Vision*, Barcelona, Spain, November 6–13 2011, pp. 1623–1630.
34. G. Riegler, M. Rüther and H. Bischof, Atgv-net: Accurate depth super-resolution, *Computer Vision-ECCV 2016-14th European Conf. Proc. Part III*, Amsterdam, The Netherlands, October 11–14 2016, pp. 268–284.
35. S. Ross, D. Munoz, M. Hebert and J. A. Bagnell, Learning message-passing inference machines for structured prediction, *The 24th IEEE Conf. Computer Vision and Pattern Recognition*, Colorado Springs, CO, USA, June 20–25 2011, pp. 2737–2744.

36. W. Shi, J. Caballero, F. Huszar, J. Totz, A. P. Aitken, R. Bishop, D. Rueckert and Z. Wang, Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network, *2016 IEEE Conf. Computer Vision and Pattern Recognition*, Las Vegas, NV, USA, June 27–30 2016, pp. 1874–1883.
37. T. Tong, G. Li, X. Liu and Q. Gao, Image super-resolution using dense skip connections, *IEEE Int. Conf. Computer Vision*, Venice, Italy, October 22–29 2017, pp. 4809–4817.
38. Z. Tu and X. Bai, Auto-context and its application to high-level vision tasks and 3d brain image segmentation, *IEEE Trans. Pattern Anal. Mach. Intell.* **32**(10) (2010) 1744–1757.
39. Z. Wang, D. Liu, J. Yang, W. Han and T. S. Huang, Deep networks for image super-resolution with sparse prior, *2015 IEEE Int. Conf. Computer Vision*, Santiago, Chile, December 7–13 2015, pp. 370–378.
40. Y. Wen, B. Sheng, P. Li, W. Lin and D. D. Feng, Deep color guided coarse-to-fine convolutional network cascade for depth image super-resolution, *IEEE Trans. Image Process.* **28**(2) (2019) 994–1006.
41. J. Xie, R. S. Feris and M. Sun, Edge-guided single depth image super resolution, *IEEE Trans. Image Process.* **25**(1) (2016) 428–438.
42. F. Xu, Z. Yan, G. Xiao, K. Zhang and W. Zuo, JPEG image super-resolution via deep residual network, *Intelligent Computing Methodologies-14th Int. Conf. Proc. Part III*, Wuhan, China, August 15–18 2018, pp. 472–483.
43. J. Yang, X. Ye, K. Li, C. Hou and Y. Wang, Color-guided depth recovery from RGB-D data using an adaptive autoregressive model, *IEEE Trans. Image Process.* **23**(8) (2014) 3443–3458.
44. X. Ye, X. Duan and H. Li, Depth super-resolution with deep edge-inference network and edge-guided depth filling, *2018 IEEE Int. Conf. Acoustics, Speech and Signal Processing*, Calgary, AB, Canada, April 15–20 2018, pp. 1398–1402.
45. A. R. Zamir, T. Wu, L. Sun, W. B. Shen, B. E. Shi, J. Malik and S. Savarese, Feedback networks, *2017 IEEE Conf. Computer Vision and Pattern Recognition*, Honolulu, HI, USA, July 21–26 2017, pp. 1808–1817.
46. K. Zhang, W. Zuo, S. Gu and L. Zhang, Learning deep cnn denoiser prior for image restoration, *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, Honolulu, HI, USA, July 21–26 2017, pp. 3929–3938.
47. Y. Zhang, Y. Tian, Y. Kong, B. Zhong and Y. Fu, Residual dense network for image super-resolution, *2018 IEEE Conf. Computer Vision and Pattern Recognition*, Salt Lake City, UT, USA, June 18–22 2018, pp. 2472–2481.