PMBANet: Progressive Multi-Branch Aggregation Network for Scene Depth Super-Resolution

Xinchen Ye , Baoli Sun, Zhihui Wang, Jingyu Yang, Rui Xu , Haojie Li, and Baopu Li

Abstract

Depth map super-resolution is an ill-posed inverse problem with many challenges. First, depth boundaries are generally hard to reconstruct particularly at large magnification factors. Second, depth regions on fine structures and tiny objects in the scene are destroyed seriously by downsampling degradation. To tackle these difficulties, we propose a progressive multi-branch aggregation network (PMBANet), which consists of stacked MBA blocks to fully address the above problems and progressively recover the degraded depth map. Specifically, each MBA block has multiple parallel branches: 1) The reconstruction branch is proposed based on the designed attention-based error feed-forward/-back modules, which iteratively exploits and compensates the downsampling errors to refine the depth map by imposing the attention mechanism on the module to gradually highlight the informative features at depth boundaries. 2) We formulate a separate guidance branch as prior knowledge to help to recover the depth details, in which the multi-scale branch is to learn a multi-scale representation that pays close attention at objects of different scales, while the color branch regularizes the depth map by using auxiliary color information. Then, a fusion block is introduced to adaptively fuse and select the discriminative features from all the branches. The design methodology of our whole network is well-founded, and extensive experiments on benchmark datasets demonstrate that our method achieves superior performance in comparison with the state-of-the-art methods. Our code and models are available at https://github.com/Sunbaoli/PMBANet_DSR/.

1 Motivation

2 Method

Traditionally, color-guided depth SR is formulated as an optimization problem, which includes a fidelity term and a prior term to make the ill-posed problem well constrained. It can be roughly summarized into the following optimization function:[1]

$$x^* = \arg\min_x \frac{1}{2}\|y - Kx\|^2 + \lambda \sum_l \omega_l * \rho_l\left(f_l \otimes x\right) \tag{1}$$

where K is a downsampling degradation matrix, λ is a trade-off parameter. f_l and $\rho_l(\cdot)$ are a set of filtering kernels and penalty functions, respectively. ω is the weighting matrix computed from the corresponding color image. $\otimes$ is the convolution operator.

3 Experiments

During training, we use 36 RGB-D images (6, 21, 9 images from 2001, 2006 and 2014 datasets, respectively) from Middlebury dataset. To evaluate the perfor- mance of our PMBANet, we test on 6 standard depth maps (Art, Books, Moebius, Dolls, Laundry, Reindeer) from Middlebury 2005, 4 standard depth maps (Tsukuba, Venus, Teddy, Cones) from Middlebury 2003.

We augment the training dataset by 180-rotation and randomly extracted 10000+ depth patches of a fixed size of 16×16 from LR depth maps. The corresponding HR depth patches are the squared size of 32, 64, 128, and 256 according to 2, 4, 8, and 16 up-scaling factors respectively. Similar to other works, the metric of Mean Absolute Difference (MAD), Root Mean Square Error (RMSE), and percentage of error pixels (PE) is used to measure the difference between the predicted depth map and the corresponding ground truth.

Lower MAD and PE values, bettern performance.

[1] A key point for these methods is to design any component in the prior term.

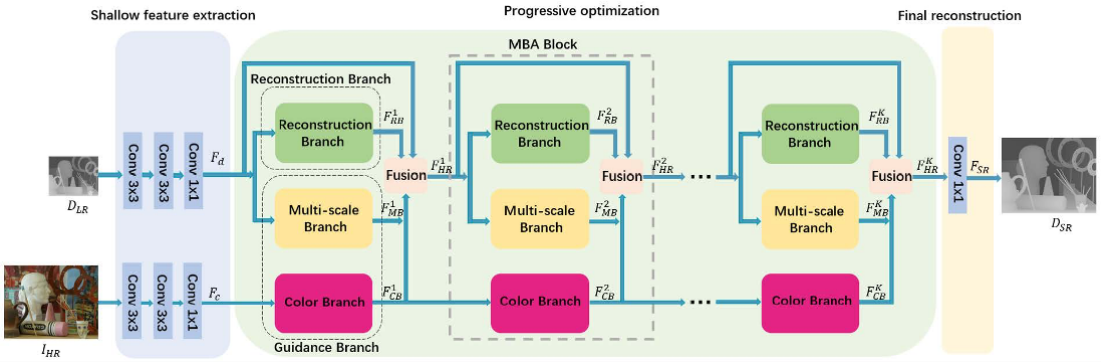

Figure 1: Network architecture of the proposed PMBANet. To better present the whole framework and implementation details, different colored rectangles are used to represent different stages and different operations in each stage.

Table 1: Quantitative depth SR results (in MAD) on Middlebury 2005 dataset.

	Art			Books			Dolls			Laundry			Mobius			Reindeer		
	×4	×8	×16	×4	×8	×16	×4	×8	×16	×4	×8	×16	×4	×8	×16	×4	×8	×16
CLMF	0.76/8.12	1.44/17.28	2.87/33.25	0.28/3.27	0.51/7.25	1.02/16.09	0.34/4.40	0.60/8.76	1.01/18.32	0.50/5.50	0.80/12.67	1.67/25.40	0.29/4.13	0.51/8.42	0.97/17.27	0.51/4.65	0.84/9.96	1.55/18.34
JGF	0.47/3.25	0.78/7.39	1.54/14.31	0.24/2.14	0.43/5.41	0.81/12.05	0.33/3.23	0.59/7.29	1.06/15.87	0.36/2.60	0.64/4.54	1.20/8.69	0.25/3.36	0.46/6.45	0.80/12.33	0.38/2.27	0.64/5.17	1.09/11.84
EDGE	0.65/6.82	1.03/13.49	2.11/25.90	0.30/3.35	0.56/8.50	1.03/19.32	0.31/2.90	0.56/6.84	1.05/17.97	0.32/2.82	0.54/5.46	1.14/13.57	0.29/3.72	0.51/7.36	1.10/14.05	0.37/2.67	0.63/6.22	1.28/16.80
TGV	0.65/5.14	1.17/10.51	2.30/21.37	0.27/2.48	0.42/4.65	0.82/11.20	0.33/4.45	0.70/11.12	2.20/45.54	0.55/6.99	1.22/16.32	3.37/53.61	0.29/3.68	0.49/6.84	0.90/14.09	0.49/4.67	1.03/11.22	3.05/43.48
KSVD	0.64/3.46	0.81/5.18	1.47/8.39	0.23/2.13	0.52/3.97	0.76/8.76	0.34/4.53	0.56/6.18	0.82/12.98	0.35/2.19	0.52/3.89	1.08/8.79	0.28/2.08	0.48/4.86	0.81/8.97	0.47/2.19	0.57/5.76	0.99/12.67
CDLLC	0.53/2.86	0.76/4.59	1.41/7.53	0.19/1.34	0.46/3.67	0.75/8.12	0.31/4.61	0.53/5.94	0.79/12.64	0.30/2.08	0.48/3.77	0.96/8.25	0.27/1.98	0.46/4.59	0.79/7.89	0.43/2.09	0.55/5.39	0.98/11.49
PB	0.79/3.12	0.93/6.18	1.98/12.34	0.16/1.39	0.43/3.34	0.79/8.12	0.53/3.99	0.83/6.22	0.99/12.86	1.13/2.68	1.89/5.62	2.87/11.76	0.17/1.95	0.47/4.12	0.82/8.32	0.56/6.04	0.97/12.17	1.89/21.35
EG	0.48/2.48	0.71/3.31	1.35/5.88	0.15/1.23	0.36/3.09	0.70/7.58	0.27/2.72	0.49/5.59	0.74/12.06	0.28/1.62	0.45/2.86	0.92/7.87	0.23/1.88	0.42/4.29	0.75/7.63	0.36/1.97	0.51/4.31	0.95/9.27
SRCNN	0.63/7.61	1.21/14.54	2.34/23.65	0.25/2.88	0.52/7.98	0.97/15.24	0.29/3.93	0.58/8.34	1.03/16.13	0.40/6.25	0.87/13.63	1.74/24.84	0.25/3.63	0.43/7.28	0.87/14.53	0.35/3.84	0.75/7.98	1.47/14.78
DSP	0.73/7.83	1.56/15.21	3.03/31.32	0.28/3.19	0.61/8.52	1.31/16.73	0.32/4.74	0.65/9.53	1.45/19.37	0.45/6.19	0.98/12.86	2.01/22.96	0.31/3.89	0.59/8.23	1.26/16.58	0.42/3.59	0.84/7.23	1.73/14.12
ATGVNet	0.65/3.78	0.81/3.78	1.42/9.68	0.43/5.48	0.51/7.16	0.79/10.32	0.41/4.55	0.52/6.27	0.56/12.64	0.37/2.07	0.89/3.78	0.94/8.69	0.38/3.47	0.45/4.81	0.80/8.56	0.41/3.82	0.58/5.68	1.01/12.63
MSG	0.46/2.31	0.76/4.31	1.53/8.78	0.15/1.21	0.41/3.24	0.76/7.85	0.25/2.39	0.51/4.86	0.87/9.94	0.30/1.68	0.46/2.78	1.12/7.62	0.21/1.79	0.43/4.05	0.76/7.48	0.31/1.73	0.52/2.93	0.99/7.63
DGDIE	0.48/2.34	1.20/13.18	2.44/26.32	0.30/3.21	0.58/7.33	1.02/14.25	0.34/4.79	0.63/9.44	0.93/11.66	0.35/2.03	0.86/3.69	1.56/16.72	0.28/1.98	0.58/8.11	0.98/16.22	0.35/1.76	0.73/7.82	1.29/15.83
DEIN	0.40/2.17	0.64/3.62	1.34/6.69	0.22/1.68	0.37/3.20	0.78/8.05	0.22/1.73	0.38/3.38	0.73/9.95	0.23/1.70	0.36/3.27	0.81/7.71	0.20/1.89	0.35/3.02	0.73/7.42	0.26/1.40	0.40/2.76	0.80/5.88
CCFN	0.43/2.23	0.72/3.59	1.50/7.28	0.17/1.19	0.36/3.07	0.69/7.32	0.25/1.98	0.46/4.49	0.75/9.84	0.24/1.39	0.41/2.49	0.71/7.35	0.23/2.18	0.39/3.91	0.73/7.41	0.29/1.51	0.46/2.79	0.95/6.58
GSRPT	0.48/2.53	0.74/4.18	1.48/7.83	0.21/1.77	0.38/4.23	0.38/4.23	0.28/2.84	0.48/4.61	0.79/10.12	0.33/1.79	0.56/4.55	1.24/8.98	0.24/2.02	0.49/4.70	0.80/8.38	0.31/1.58	0.61/5.90	1.07/10.35
Ours	0.26/1.95	0.51/3.45	1.22/6.28	0.15/1.13	0.26/2.87	0.59/6.79	0.19/1.35	0.32/3.22	0.59/8.92	0.17/1.27	0.34/2.41	0.71/6.88	0.16/1.21	0.26/2.87	0.67/6.73	0.17/1.28	0.34/2.40	0.74/5.66

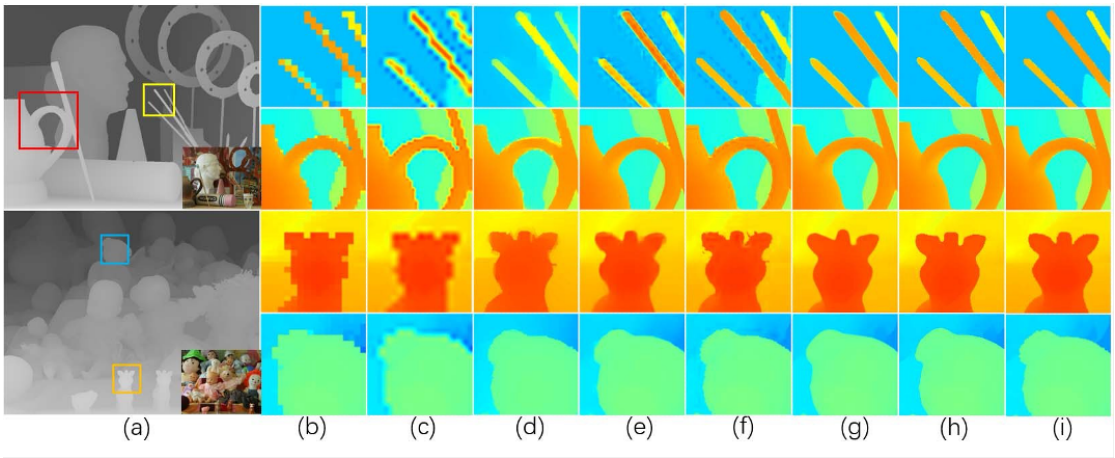

Figure 2: Visual comparison of 8× upsampling results on two examples (Art, Dolls). (a) GT depth maps and color images; (b) LR; (c) Bicubic; (d) JGF; (e) DGDIE; (f) DEIN; (g) GSRPT; (h) PMBANet; (i) GT. Depth patches are enlarged and colored to enhance the contrast for clear visualization.