High-Order Residual Network for Light Field Super-Resolution

AAAI-20

[1]Nan Meng, [2]Xiaofei Wu, [2]Jianzhuang Liu, [1]Edmund Y. Lam
[1]Depart of Electrical and Electronic Engineering, University of Hong Kong
[2]Huawei Noah's Ark Lab, China

Introduction

1. We propose to establish a framework tailored to the light field (LF) structural information.
2. We propose a novel high-order structure, named high-order residual block (HRB) to learn the features by fully considering the information from all sub-aperture images of a LF. Such features extracted from HRB preserve high angular coherence.
3. By stacking a set of the HRBs, the proposed network is able to extract diverse spatial features endowed with scene geometry information.
4. Experimental results demonstrate that our model not only outperforms the state-of-the-art reconstruction methods on quantitative measurements and visual quality.

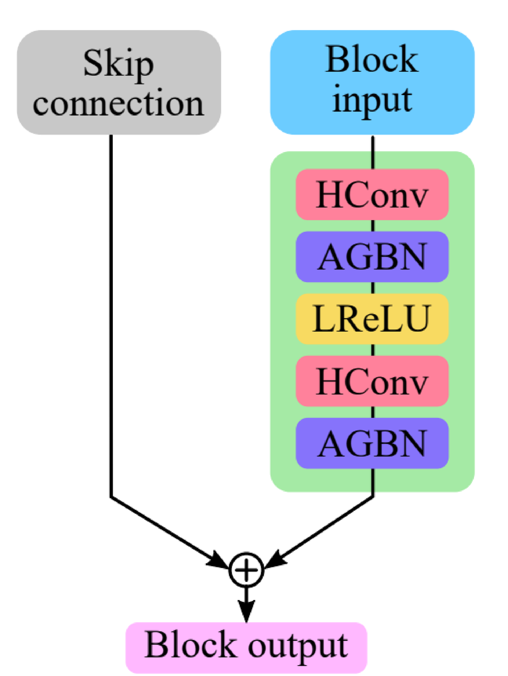
Problem formulation

The LF reconstruction deals with the recovery of the high-resolution LF $I^H(x,y,u,v)$ from the corresponding low-resolution LF $I^L(x,y,u,v)$.

$$I^S(x,y,u,v) = g(I^L(x,y,u,v), \Theta),$$

where $g(\cdot)$ denotes the mapping from LR to super-resolved LF.

HRB module



1. Each HRB contains two HConv layers with the 3×3 angular receptive field, allowing it to fully utilize the information from all SAIs of input features.
2. We propose the aperture group batch normalization (AGBN) to avoid that the whitening decorrelates the coherence among feature SAIs.

Fig.1 Structure of HRB.

Network

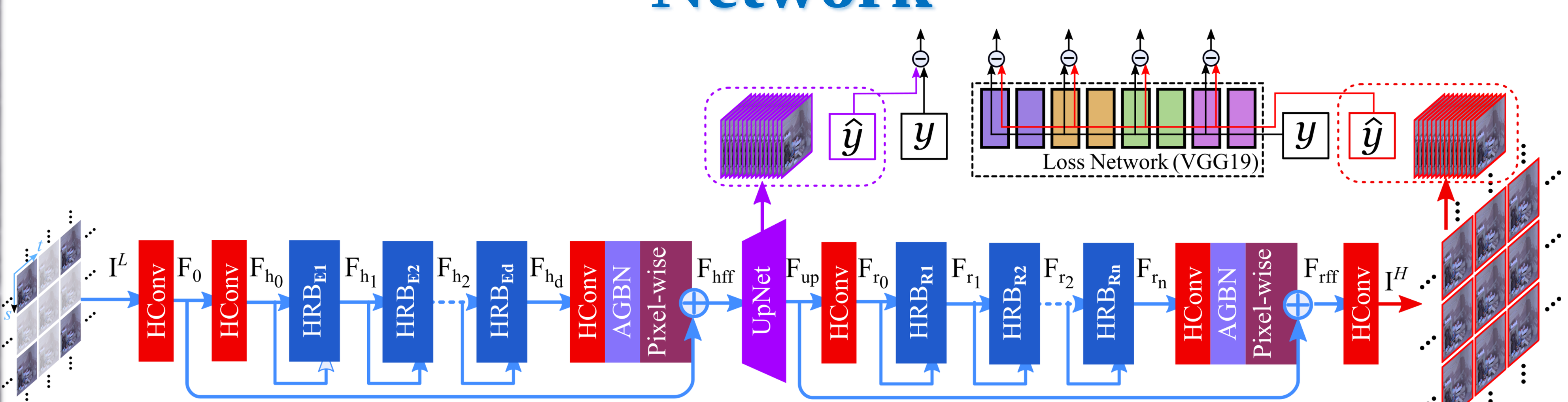


Fig.2 The architecture of our hierarchical high-order network

Network Modules

1. Shallow Feature Extraction:

$$F_0 = H_{HC}(I^L)$$

2. GRLNet:

$$F_{G_d} = H^d_{\mathrm{HRB}}\left(F_{G_{d-1}}\right) = H^d_{\mathrm{HRB}}\left(H^{d-1}_{\mathrm{HRB}}(F_{G_{d-2}})\right) = H^d_{\mathrm{HRB}} \circ H^{d-1}_{\mathrm{HRB}} \circ \cdots \circ H^1_{\mathrm{HRB}}(F_{G_0})$$

$$F_G = H_{AGBN} \circ H_{HC}(F_{G_d}) + F_0$$

3. UpNet:

$$F_{up} = H_{up}(F_G)$$

4. SReNet:

$$F_R = H_{AGBN} \circ H_{HC}(F_{R_n}) + F_{up}$$

$$F_{R_n} = H^n_{\mathrm{HRB}} \circ H^{n-1}_{\mathrm{HRB}} \circ \cdots \circ H^1_{\mathrm{HRB}}(F_{R_0})$$

The proposed approach can learn the completely spatial-angular structure and redundancy, which can be useful for HD visual signal reconstruction in computational imaging.

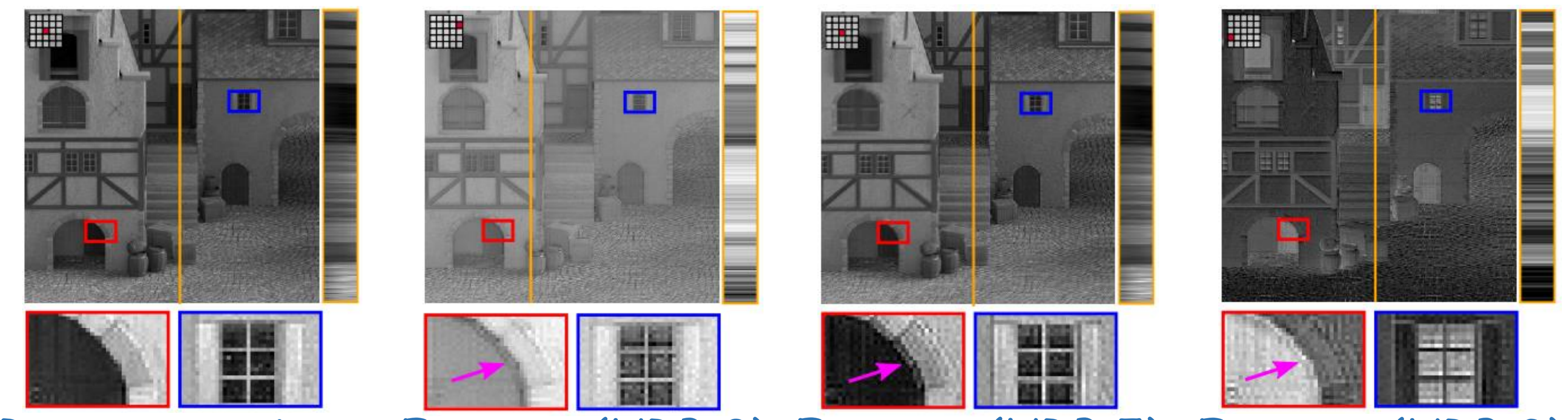


Fig.3 Geometric features extracted from different HRBs.

Loss function

We define a spatio-angular loss function to supervise the training process. Such loss is formulated as the weighted sum of a perceptual loss ℓ_P, a reconstruction loss ℓ_R (based on MSE).

$$\ell = \alpha \cdot \ell_R + \beta \cdot \ell_P,$$

where α and $\beta$are scalars.

- Perceptual loss ℓ_P: $\ell_P = \frac{1}{ST}\sum_S \sum_T (\phi(I^H_{s,t}) - \phi(g(I^L_{s,t};\Theta)))^2$
- Reconstruction loss ℓ_R: $\ell_R = \sum_X \sum_Y \sum_S \sum_T (I^H - I^S)^2$

Results

Table 1. Quantitative evaluation on PSNR and SSIM on both synthetic LF and real-world LF for Spatial $4\times$ and Angular $3\times$.

Algorithm	Scale	PSNR(dB) Occlusions (20)	PSNR(dB) Reflective (20)	PSNR(dB) HCI new (8)	PSNR(dB) HCI old	PSNR(dB) EPFL (21)	SSIM Occlusions (20)	SSIM Reflective (20)	SSIM HCI new (8)	SSIM HCI old	SSIM EPFL (21)
Bicubic	S×4	24.98	27.54	25.92	23.95	25.94	0.663	0.771	0.688	0.630	0.767
LFCNN		25.04	28.14	28.28	25.65	26.97	0.686	0.798	0.768	0.688	0.792
BM PCA+RR		26.28	28.73	28.90	25.85	27.51	0.710	0.796	0.772	0.703	0.785
LFNet		25.94	28.81	29.36	25.40	26.10	0.709	0.808	0.762	0.706	0.775
Zhang et al.		27.36	29.69	28.74	25.83	28.02	0.744	0.810	0.743	0.694	0.798
MSLapSRN		27.41	30.28	29.55	26.27	28.78	0.755	0.835	0.782	0.723	0.821
RDN		26.97	29.64	29.63	26.66	28.58	0.724	0.817	0.792	0.730	0.804
Ours		**29.04**	**30.82**	**30.78**	**28.54**	**29.63**	**0.815**	**0.859**	**0.837**	**0.821**	**0.859**
Kalantari et al.	A×3	34.70	37.24	35.53	32.59	35.19	0.927	0.958	0.916	0.906	0.959
Wu et al.		35.64	40.03	35.64	33.38	37.05	0.928	0.963	**0.918**	0.905	0.960
Ours (4HRBs)		**35.96**	**40.16**	**35.79**	**33.75**	**38.28**	**0.930**	**0.964**	0.904	**0.913**	**0.961**

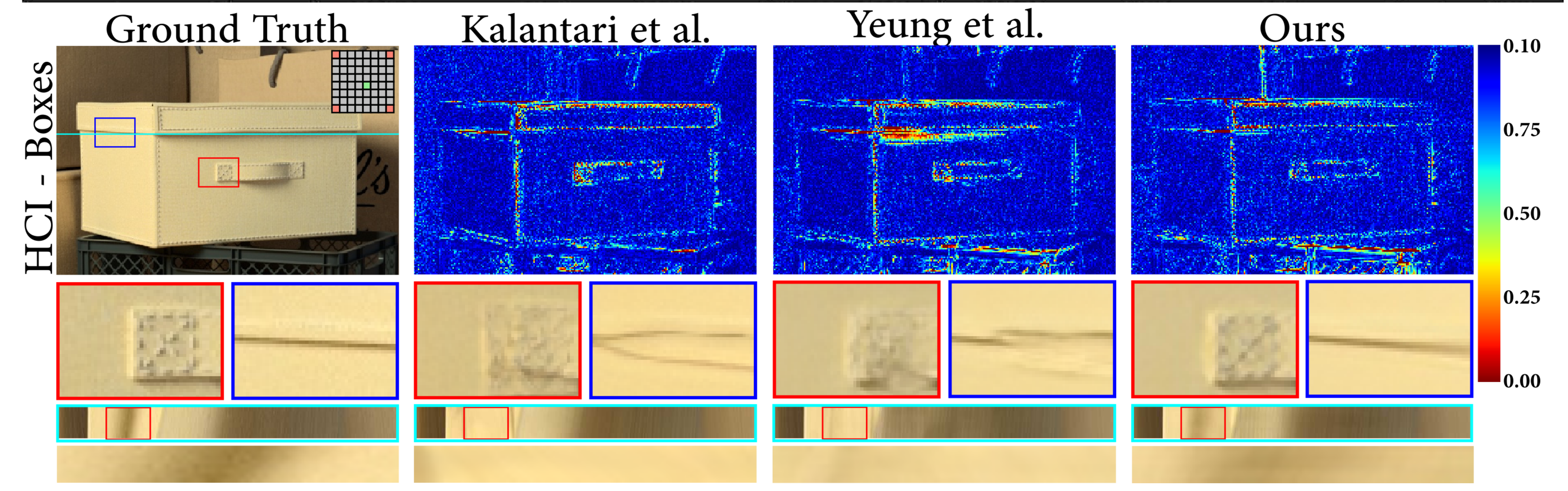


Fig 4. Visual comparisons on a real-world scene (HCI - Boxes) for angular $2\times 2 \longrightarrow 8\times 8$ SR.

1. We compare the quantitative results with both SISR methods and LFSR methods on 5 public LF datasets.
2. The proposed model outperforms existing SOTA methods on both real-world and synthetic scenes.
3. The visual results show that our model is able to generate the spatial details with better image fidelity compared with SOTA methods.

Related Work

[1] N. Meng, H. K.-H. So, X. Sun, and E. Y. Lam, "High-dimensional dense residual convolutional neural network for light field reconstruction," IEEE Trans on Pattern Analysis Machine Intelligence.