Revisiting Light Field Rendering with Deep Anti-Aliasing Neural Network (Supplementary Material)

Gaochang Wu, Yebin Liu, *Member, IEEE,* Lu Fang, *Senior Member, IEEE,* and Tianyou Chai, *Fellow, IEEE*

In Sec. I, we provide an additional Fourier analysis of the downscaling operation compared to the conventional reconstruction filter. Combining with the analysis in the main paper, we can draw a conclusion that the proposed "shearing-downscaling-prefiltering" framework in image domain is comparable to the conventional reconstruction filter in the Fourier domain. In Sec. II, we provide additional reconstruction results, including four qualitative results, more quantitative evaluations and two 4D error maps. In Sec. III, we discuss the prefiltering and shearing operations after being implemented in the end-to-end optimized network, and describe the implementation details of the view extrapolation.

I. Fourier Analysis of the Downscaling Operation

In this section, we will further analyse the "downscaling operation" in the Fourier domain in both ideal and practical implementations.

In the following, we will first theoretically analyse the downscaling operation under general conditions. Our original idea behind the utilizing of "downscaling operation" is to increase the sample interval (e.g., from Δu to $\Delta u'$) along the spatial dimension [5], which can be interpreted as reducing camera resolution [5, 26] or disparity [7, 8], as mentioned in the manuscript. With a proper sample interval, the signal will be bounded by the bandlimit $\frac{\pi}{\Delta'_u}$[1]. As illustrated in Fig. 1(a), increasing sample interval will directly block the aliasing high frequencies. On the other side, the conventional reconstruction filter can be interpreted as a low-pass filter to reduce the highest frequencies in the spectra support, which is equivalent to increasing the sample interval when considering an ideal low-pass filter. As illustrated in Fig. 1*(b), the ideal downscaling (increasing sample interval) and the ideal conventional reconstruction filter (ideal low-pass filtering) produce results with similar qualities. We can safely draw a conclusion that the downscaling operation generally equals to the conventional reconstruction filter under ideal conditions.

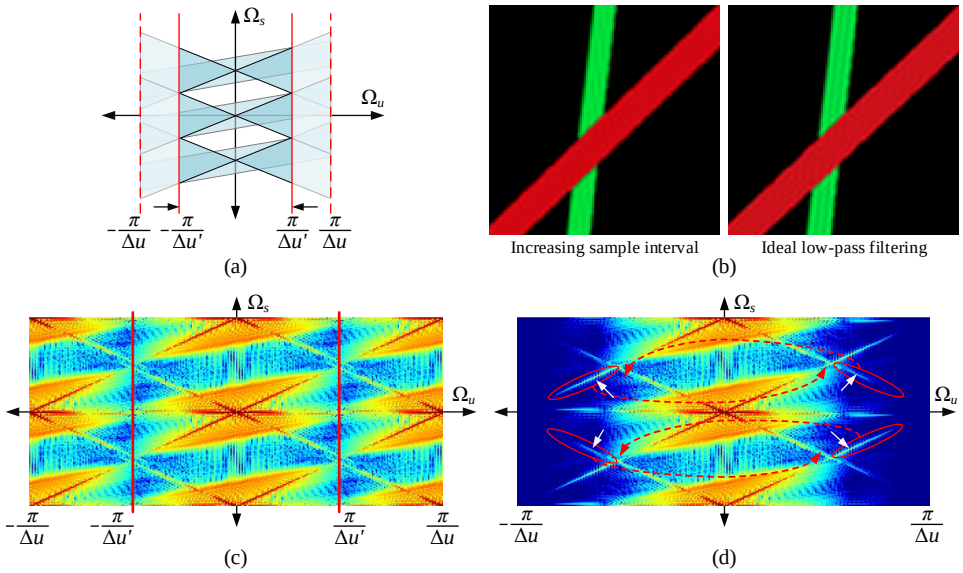


Fig. 1: Fourier analysis of the downscaling operation. (a) An ideal downscaling operation increases the sample interval (from Δu to $\Delta u'$) along the spatial dimension, which blocks the aliasing high frequencies in the Fourier domain; (b) Similar results can be produced by increasing sample interval and ideal low-pass filtering; (c) Practical downscaling employs an anti-aliasing interpolation method [32], which barely introduce aliasing components to the spectral support within region $[-\frac{\pi}{\Delta u'}, \frac{\pi}{\Delta u'}]$; (d) Some aliasing residuals still exist in the spectra support after the conventional reconstruction filtering, as shown by the white arrows.

[1]As indicated in [5], "the maximum camera spacing will be larger if the scene texture variation gets more uniform, or if the rendering camera resolution becomes lower." Reducing camera (rendering camera is regarded as sampling camera in [5]) resolution means increasing sample interval along the spatial dimension, e.g., from Δu to $\Delta u'$.

In practical implementation, the downscaling operation is also comparable to the conventional reconstruction filter. In practice, both the downscaling operation and conventional reconstruction filtering are implemented using a low-pass filter, the former also includes a downsampling. On the one hand, practical downscaling operation employs an anti-aliasing interpolation method [32], whose kernel size is in proportion to the downsampling factor. The aliasing residuals after the downsampling operation have little effect on the original (target) spectral support within region $[-\frac{\pi}{\Delta u'}, \frac{\pi}{\Delta u'}]$, as shown in Fig. 1(c). On the other hand, after the practical low-pass filtering, there are still some residuals in the spectra support, as shown by the white arrows in Fig. 1(d). We can block the signal within region $[-\frac{\pi}{\Delta u'}, \frac{\pi}{\Delta u'}]$. This operation is technically equivalent to the downsampling. The downsampling operation copies the aliasing residuals with offset $\frac{2\pi}{\Delta u'}$ (or $-\frac{2\pi}{\Delta u'}$), as shown by the red arrows in Fig. 1(d). Since the amplitude of the aliasing residuals is attenuated, the offset of the aliasing residuals can hardly influence the reconstructed result.

In conclusion, the proposed "shearing-downscaling-prefiltering" framework in image domain is comparable to the conventional reconstruction filter in the Fourier domain.

II. Additional Reconstruction Results

Fig. 2 shows additional results on light fields from the ICME 2018 Grand Challenge on Densely Sampled Light Field Reconstruction [40] ("ICME DSLF dataset" for short).

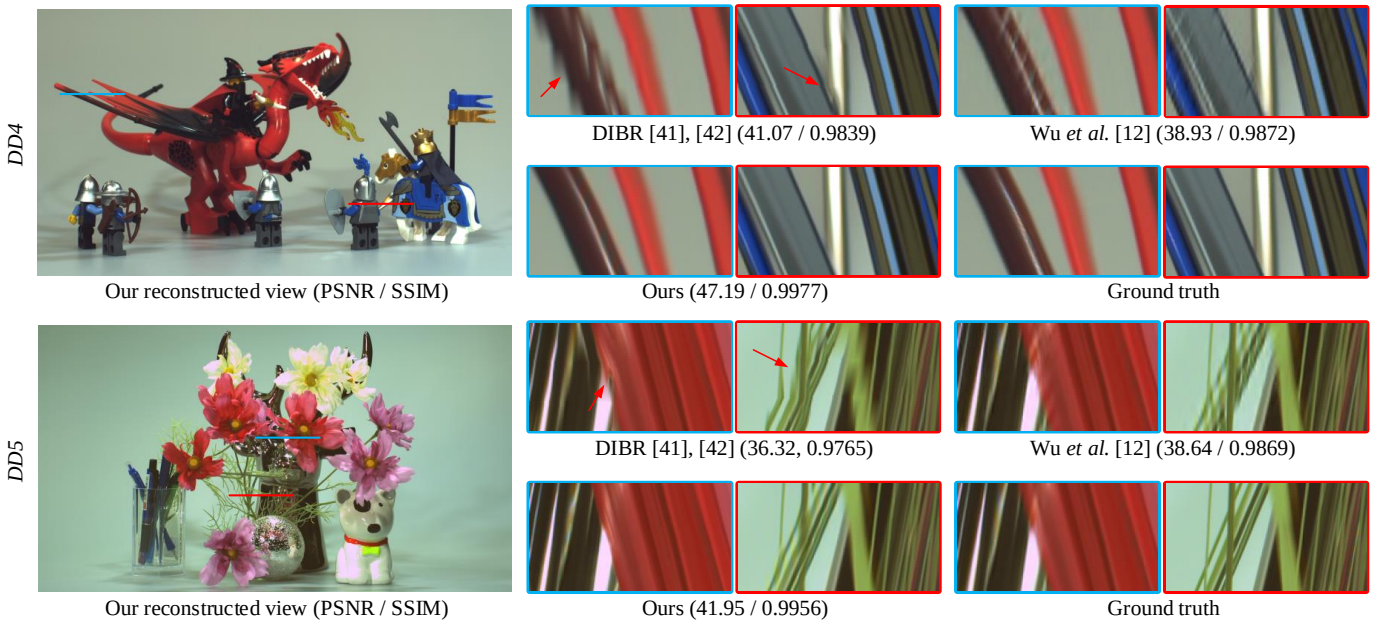


Fig. 2: Comparison of the results (16× upsampling) on the light fields from the ICME DSLF dataset [40]. The PSNR and SSIM values are averaged on the reconstructed views (input views are excluded).

Table I lists additional quantitative evaluations of $\times 24$ and $\times 32$ upsampling rates on the ICME DSLF dataset [40] (*DD1*, *DD2* and *DD3*). It can be clearly shown that the proposed DA^{2}N can handle relatively large disparity range (21px, 28px etc.), without depth estimation.

TABLE I: Quantitative results (PSNR/SSIM) of reconstructed LFs on the LFs from the ICME DSLF dataset [34].

Method	*DD1*		*DD2*		*DD3*		Average
	(21px) ×24	(28px) ×32	(21px) ×24	(28px) ×32	(21px) ×24	(28px) ×32	
DIBR [41], [42]	38.56/0.9857	37.38/0.9834	37.01/0.9647	35.15/0.9610	34.31/0.9689	31.70/0.9593	35.69/0.9705
Wu *et al.* [12]	38.71/0.9763	36.52/0.9652	33.77/0.9720	31.95/0.9639	30.80/0.9573	28.25/0.9295	33.33/0.9607
Our proposed	**45.09/0.9941**	**43.16/0.9909**	**37.52/0.9855**	**36.53/0.9827**	**35.07/0.9834**	**33.28/0.9755**	**38.44/0.9854**

The disparity range ($d_{\max} - d_{\min}$) between the input neighboring views is listed in the table.

Fig. 3 shows additional results on light fields from Lytro Illume [49]. The method without explicit depth by Yeung *et al.* [13] and the proposed network work better on those non-Lambertian regions. In addition, we provide a detailed line chart for each category of light field from Lytro Illum comparing with the state-of-the-art method by Yeung et al. [13] in Fig. 4.

Fig. 5 demonstrates the 4D view consistency by projecting the 4D error maps onto the 2D plane. The demonstrate cases show 7×7 reconstruction results using 3×3 inputs. In each sub-figure, we demonstrate 4D error map (top), 2D error map at unsampled angular line (bottom left) and 2D RMSE map for each view (bottom right). It can clearly shown that our scheme

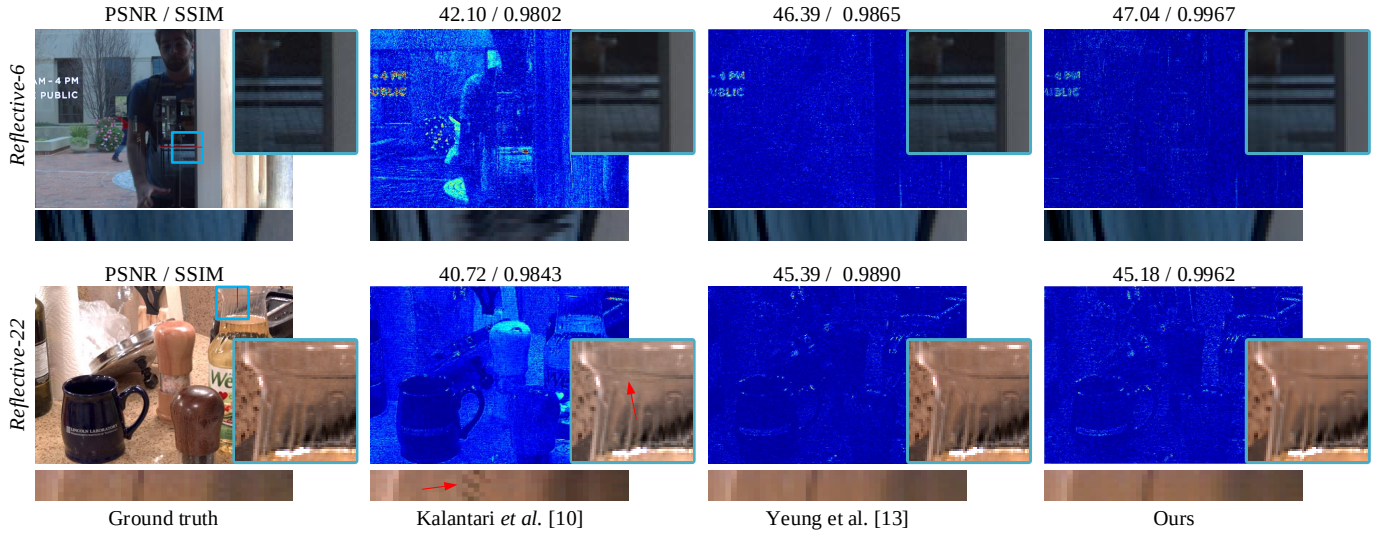
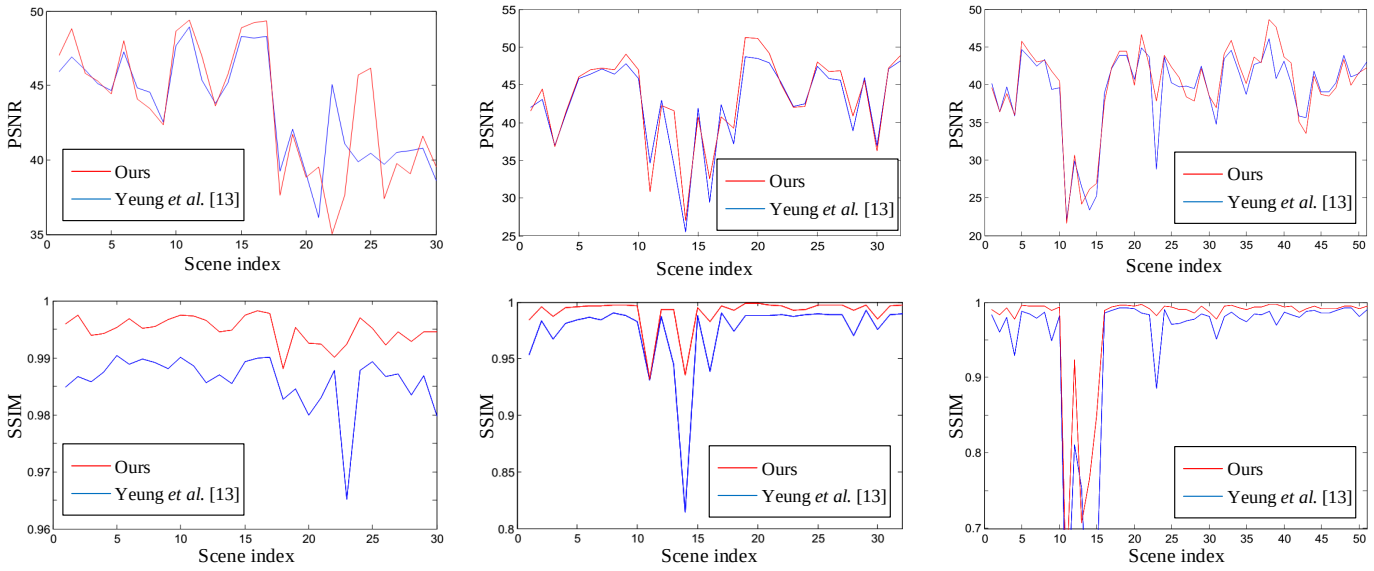


Fig. 3: Comparison of the results on the LFs from Lytro Illume (×3 upsampling). The proposed network and the method by Yeung *et al.* [13] are able to reconstruct the non-Lambertian effects.

leads to relatively smaller errors between views at unsampled angular line and those at sampled angular line, compared with other methods (Kalantari *et al*. [10] and Yeung *et al*. [13]). It indicates that the proposed EPI-based method is able to retain 4D view consistency.



(a) *30 Scenes* (b) *Reflective & Refractive Surfaces* (c) *Occlusions*

Fig. 4: Detailed line chart for each category of light field from Lytro Illum for 3× upsampling.

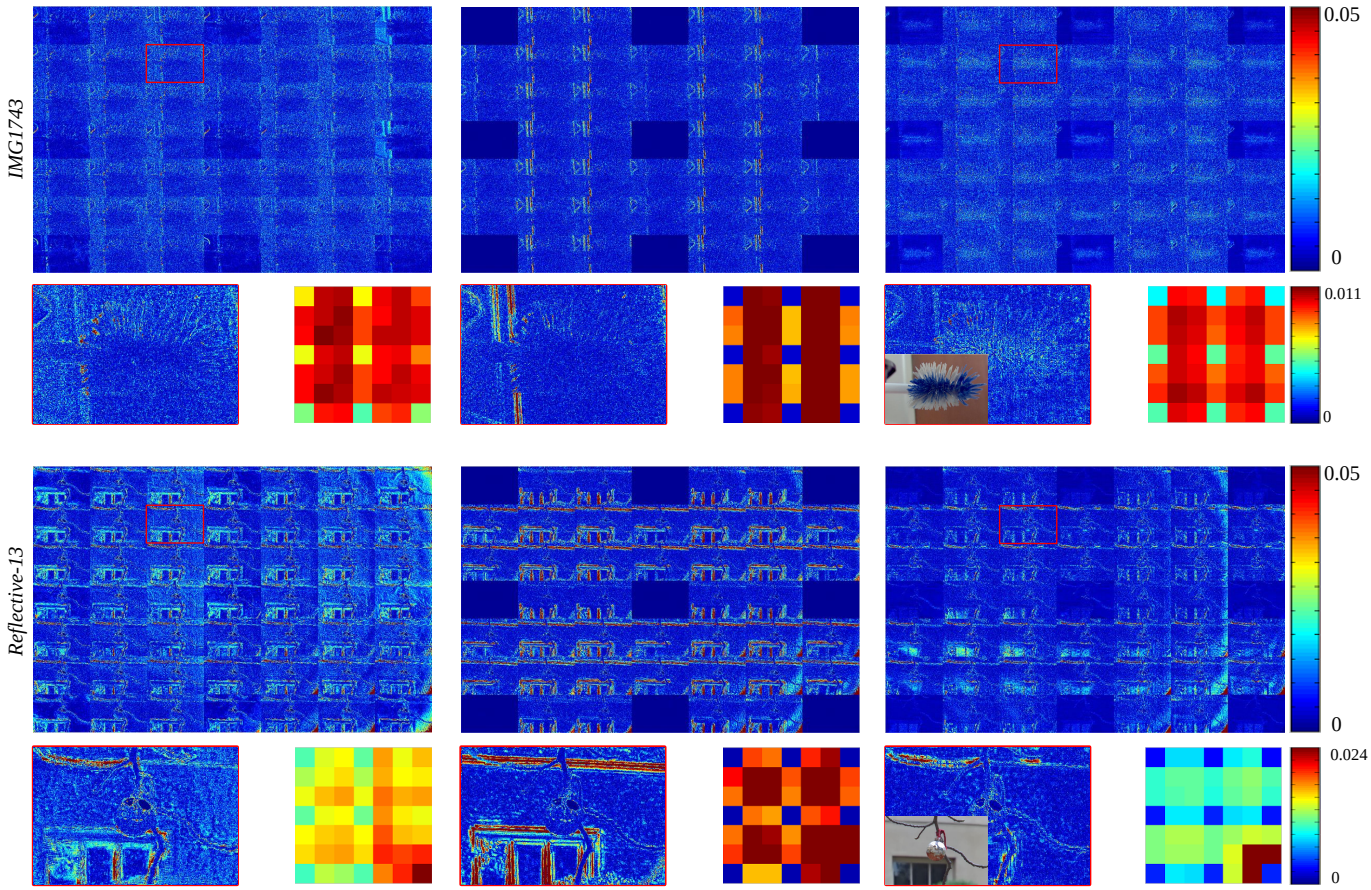
III. Additional Discussions

A. Discussion on the Prefiltering

Although we initialize the kernels in the prefiltering layer by Gaussian functions with different shape parameters σ_c, the final shapes of the filters are no longer Gaussian after the network training. We visualize the convolutional kernel of the layer "conv2_3" before (left) and after (right) the end-to-end training in Fig. 6.

B. Discussion on the Shearing

In this subsection, we will discuss the superiority of implementing shearing operation in deep network comparing with the classical reconstruction filtering and the influence under different settings of shear range.



(a) Kalantari *et al.* [10] (b) Yeung *et al.* [13] (c) Ours

Fig. 5: Reconstruction error maps (4D) for LFs from Lytro Illum.

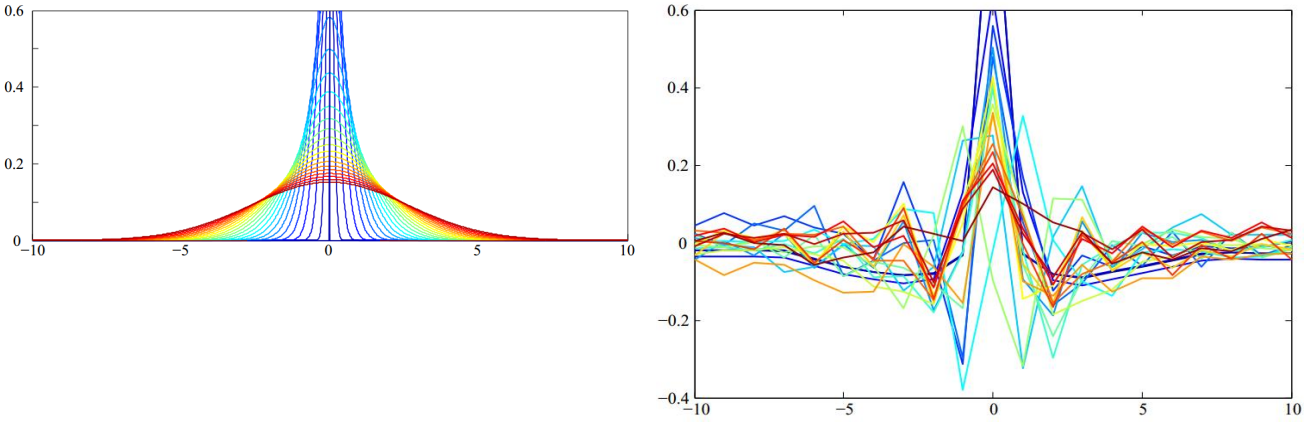
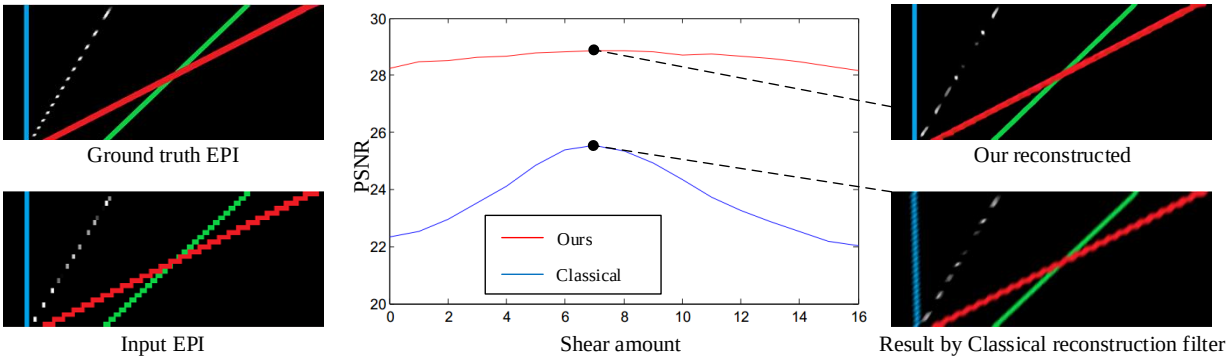


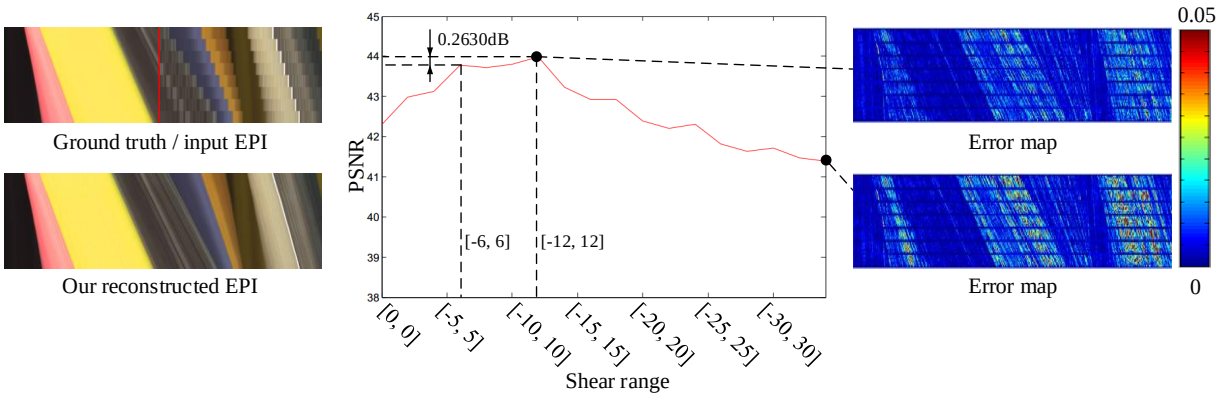
Fig. 6: The convolutional kernel in the prefiltering layer before (left) and after (right) the end-to-end training.

Fig. 7(a) demonstrates the performance of the proposed network under **different shear amounts**. For better illustration, we force the shear amount for each branch of the network to be identical, so does the rendering disparity for the classical reconstruction filter. The disparity range of such toy example is [0,14], i.e., the optimal rendering disparity is 7. Note that in practical, different shear amounts are assigned to the branches of the network. It can be clearly shown that the performance of the proposed approach is stable even when the shear amount is not equal to the optimal rendering disparity. While the classical reconstruction filter [5] is sensitive to the variation of the rendering disparity (i.e., shear amount).

Fig. 7(b) further demonstrates the performance of the proposed network under **different settings of shear range**. The shear amounts are assigned to 7 branches of the network uniformly. The light field is *DD1* from the ICME DSLF dataset [40]. A light field with 1×8 views with disparity range $[-17.7, 5.9]$ is used to reconstruct a 1×190 light field ($27\times$ upsampling). The setting of shear range varies from $[0, 0]$ to $[-34, 34]$. The reconstruction result reaches the highest PSNR value using shear range $[-12, 12]$ and the lowest PSNR value using shear range $[-34, 34]$. As the reconstruction quality tends to be stable



(a) Variation of shear amounts



(b) Variation of shear ranges

Fig. 7: Analysis of performance under different settings of (a) shear amounts and (b) shear ranges.

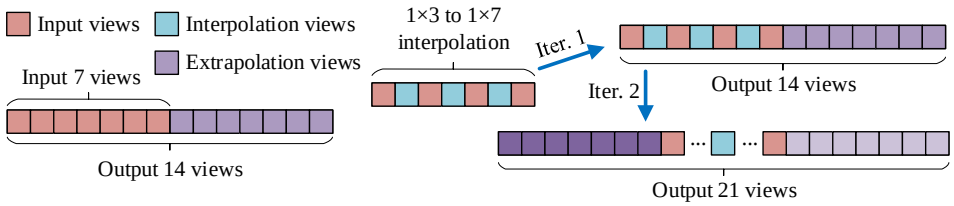
around shear range $[-9, 9]$ (from $[-6, 6]$ to $[-12, 12]$), all the results reported in the paper are produced under the setting of $[-9, 9]$.

The basic principle for determining the shear amounts is ensuring the proposed "shearing-downscaling" framework covers the disparity range of the input light field. In other words, based on the light field rendering theory that the disparity d should within 1 pixel range, the shear amount α_h is expected to satisfy $|\frac{d-\alpha_h}{\alpha_u}| \leq 1$ ($\alpha_u = 4$ is the downscaling factor), e.g., for a disparity range $[-13, 13]$ the shear range would be $[-9, 9]$. However, when considering the receptive field of a convolutional neural network, a light field with larger disparity range can also be well reconstructed.

C. Implementation Details of the View Extrapolation

In the training phase (as shown in Fig. 8(a)), we use 7 views to extrapolate a light field with 14 views, i.e., from view #1, #2, ..., #7 to view #1, #2, ..., #14. And the "deconv5" layer in the extrapolation network performs $\times 2$ upsampling.

In the reasoning (testing) phase, we first use the proposed reconstruction network to interpolate a high angular resolution light field, e.g., from 1×3 views to 1×7 views as shown in the left part of Fig. 8(b). We have mentioned in the fourth paragraph in Sec. 6.4 that "where 7 views are from the interpolation." Then a 1×14 (view #8, #9, ..., #21) light field can be obtained through extrapolation using the reconstructed 1×7 views (view #8, #9, ..., #14) in the first iteration (see the top right of Fig. 8(b)). In the second iteration, we flip the input views along the angular dimension (view #14, #13, ..., #8) as well as the spatial dimension (to ensure the occlusion relation) to eventually extrapolate a 1×21 light field (view #1, #2, ..., #21, see the bottom right of Fig. 8(b)).



(a) Training phase (b) Reasoning phase

Fig. 8: Detailed implementation of view extrapolation.