An Improved 4D Convolutional Neural Network for Light Field Reconstruction

Qiuming Liu[1,2(✉)], Ruiqin Li[1], Ke Yan[1], Yichen Wang[1], and Yong Luo[3]

[1] School of Software Engineering, Jiangxi University of Science and Technology, Nanchang 330013, China
liuqiuming@jxust.edu.cn, 6720210698@mail.jxust.edu.cn
[2] Nanchang Key laboratory of Virtual Digital Factory and Cultural Communications, Nanchang 330013, People's Republic of China
[3] School of Software, Jiangxi Normal University, Nanchang 330022, China

Abstract. Light field (LF) camera sensors often face a trade-off between angular resolution and spatial resolution when shooting. High spatial resolution image arrays often result in lower angular resolution, and vice versa. In order to obtain high spatial resolution and at the same time have high angular resolution. In this paper, we propose an improved 4D convolutional neural network (CNN) algorithm for angular super-resolution (SR) to improve the quality of angular SR images. Firstly, to address the problem of low luminance of images captured by LF cameras, this paper uses block threshold square reinforcement (BTSR) for image luminance enhancement. Secondly, to make the reconstructed new viewpoints of higher quality, this paper improves the attention mechanism convolutional block attention module (CBAM). This paper incorporates it into a 4D dense residual network as high dimensional attention module (HDAM). HDAM generates images along two independent dimensions, spatial and channel. The HDAM generates attention maps along two independent dimensions, space and channel, which guide the network to focus on more important features for adaptive feature modification. Finally, this paper modifies the activation function to make the network perform better in the later stages of training and more suitable for LF reconstruction tasks. This paper evaluates the network on many LF data, including real-world scenes and synthetic data. The experimental results show that the improved network algorithm can achieve higher quality LF reconstruction.

Keywords: Light field reconstruction · 4D convolution · Convolutional neural network · Attention mechanism

1 Introduction

Light Field (LF) is often represented as the set of all light rays in a scene, and LF cameras can be used to record 3D information about the scene. Unlike

C. Wu et al. (Eds.): MONAMI 2023, LNICST 559, pp. 108–120, 2024.
https://doi.org/10.1007/978-3-031-55471-1_9

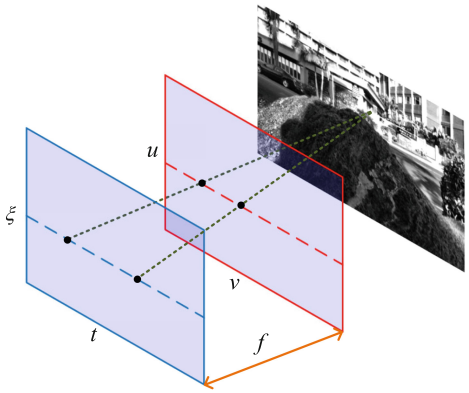

Fig. 1. 4D representation of the light field. Where (ξ, t) denotes the camera plane, (u, v) denotes the image plane, and f is the focal length of the camera.

traditional 2D imaging systems, LF cameras record the position and direction of each light $L(u, v, \xi, t)$ in the scene in two planes.This representation is shown in Fig. 1, where (ξ, t) denotes the camera plane, (u, v) denotes the imaging plane, and f is the focal length of the camera.A single exposure of the camera will result in a set of image arrays that record the light intensity of the scene as observed at different positions. Such a rich way of recording information makes many applications possible, such as depth estimation [1], refocusing [2], image segmentation [3], 3D reconstruction [4], etc. But there is a trade-off in this way of scene information capture. Because the product of spatial resolution and angular resolution cannot exceed the resolution of the sensor. A high spatial resolution necessarily makes the angular resolution lower, and vice versa.

To solve this problem, many researchers have proposed many methods to reconstruct dense LF views with sparse LF as input. The technique is called angular super-resolution (SR), also known as view synthesis. Wang *et al.* [5] used an algorithm of depth estimation to obtain an accurate depth map, and then warped the existing image into the new views. In [6], Pearson *et al.* layered the scene depth based on plenoptic function theory and rendered the new views using probabilistic interpolation. Zhang *et al.* [7] estimated the parallax information from the perspective of phase using parallax cues and phase synthesis methods, and then synthesized the new view using a parallax-based warping method. Paper [8] further developed the patched-based method, and Zhang et al. decomposed the central view into different depth layers and then performed the synthesis of the new views. On the other hand, Chai *et al.* [9] consider the rendering of new views as equivalent to the reconstruction of functions from the collected samples. They introduced plenoptic sampling into the Fourier framework for the first time by assuming an unobstructed Lambert scene, for which the effect of the maximum minimum depth on the plenoptic spectrum structure was derived. This assumption is extended to non-Lambertian scenes and obscured scenes of [10,11] in [10]. Similarly, Zhu *et al.* [12] investigated the effect of the surface curvature of the irregular geometry of the scene on the plenoptic

spectrum and used it to design an efficient reconstruction filter for the rendering of new views.Vagharshakyan *et al.* [13] treat view synthesis as a restoration task on EPI and use a sparse representation of LF in the shearlet transform for view synthesis. In [14], Chen *et al.* build a mathematical model of self-obscuration by studying the slope relationship between the parabolic tangents and the light captured by the camera. They derive a closed spectrum formulation based on the established model as a way to study the effect of occlusion on the spectrum and to derive a specific sampling rate and a new reconstruction filter. All the above methods are non-learning based methods, and the rendered new views are prone to ghosting when facing complex scenes such as reflections, occlusions, and rich textures.

In recent years, many learning-based algorithms have emerged in the field of view synthesis due to the development of deep learning. Yoon *et al.* [15] designed a deep learning algorithm that uses two adjacent views to generate a new virtual view. Flynn *et al.* [16] synthesized novel views based on image sequences with wide baselines. Wu *et al.* [17] started with EPI to obtain richer angle and parallax information. They used the "blur-restoration-deblu" framework to achieve LF reconstruction. In order to reconstruct scenes with larger parallaxes, [18] proposed a method that incorporates sheared EPIs to continue the improvement of the previous method. With the improvement of Wu *et al.*, the algorithm is able to adapt to sparse LF data with larger parallax. Despite the increased parallax, they only considered 2D EPI, and there are still some deficiencies in the fusion of angular information. Wang *et al.* [19] proposed to combine EPI and EPI volume representation of 4D LF for LF angle reconstruction. They combine 2D convolutional operations and 3D convolutional operations to construct a pseudo-4D CNN. Yeung *et al.* [20] proposed an end-to-end network for densely sampled LF reconstruction. Exploring the relationship between subaperture images (SAI) and pseudo-4D filters, the method achieves state-of-the-art performance in a large number of real scenes captured by Lytro cameras. The above methods use 2D or pseudo-4D convolutional neural network (CNN) to extract features when training the network, instead of using a real 4D CNN. The complexity of 4D LF data may lead to less comprehensive information obtained by their algorithms, and can also lead to inefficiency of the algorithms. Meng *et al.* [21] proposed to use a high-dimensional dense residual CNN to recover LF. The method takes each SAI of LF as input and captures the association relationship between the views by 4D convolution. After experiments this paper founds that the algorithm [21] still has room for improvement in terms of training preprocessing, network modules, and activation functions. Therefore, this paper improves and adjust the structure and modules of the network to make it have better performance.

It is well known that the images obtained by LF cameras such as Lytro are of low brightness. The decoded images have more bad points and more noise in the edge SAI. Directly using the decoded images for training will greatly affect the training effect of the network. To solve this problem, this paper performs block threshold square reinforcement (BTSR) on the images before angle SR.

This method equalizes the grayscale values within each block and smoothes out the extreme pixel values in the image, making the texture more visible.

In order to keep the integrity and consistency of the reconstructed image with ground truth in terms of structure and pixels, respectively. 4D CNN can achieve better SR for LF images, but it is still not sufficient to grasp the importance of semantic information in the feature map. Many researchers keep increasing the sensory field to obtain the semantic information of the scene in order to extract the image global information more comprehensively. Although the semantic information of the network becomes rich, it is not differentiated and processed separately according to the importance of the information. This not only leads to the loss of low-level image texture information, but also makes the important information regions not better processed, which in turn affects the quality of the generated new views. Therefore, this paper first improves the attention mechanism CBAM by extending its input dimension to become a high dimensional attention module (HDAM). This module can adapt to high dimensional feature maps and infer the attention map along the high dimensional feature information. By introducing an attention mechanism, it focuses on the information that is more critical to the current task among the many input information. Reducing the attention to other information and even filtering out irrelevant information. This solves the information overload problem and improves the efficiency and accuracy of task processing. In addition, this paper also replaced the activation function from LeakyReLU to GELU, which suppresses the generation of pixels with large negative values in the prediction process and gives the network a higher reconstruction effect.

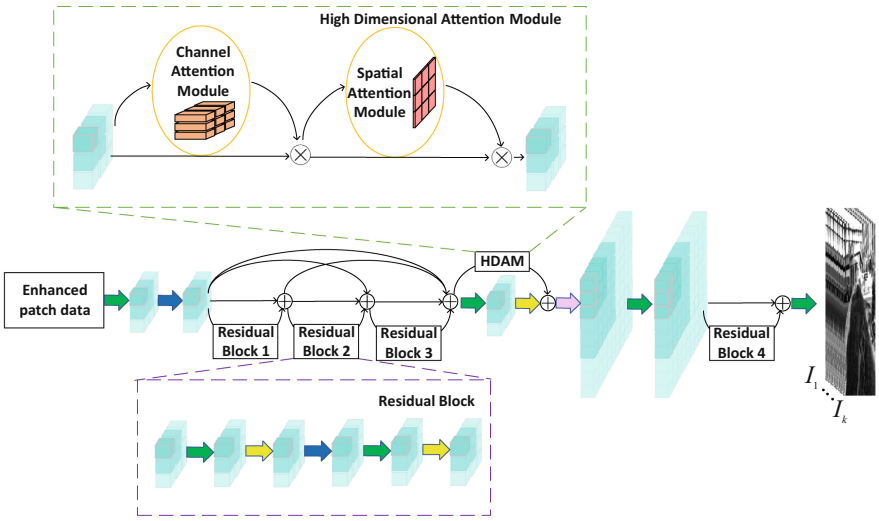

Fig. 2. The overview of the proposed model. Green arrows indicate convolution, blue arrows indicate activation functions, yellow arrows indicate normalization operations, and pink arrows indicate upsampling. (Color figure online)

In this paper, we propose an improved 4D convolutional angular SR network incorporating an attention mechanism. As shown in Fig. 2, the image is first preprocessed for enhancement, and the enhanced image needs to be converted from RGB to YCbCr format. The Y channel of the image is used as the input, and the 4D convolution is used to extract the image features, and the HDAM is used to guide the network to focus on the important regions. Finally, the feature map is upsampled to the same number of views as the ground truth to achieve angular SR.

2 Proposed Method

In this section, this paper will address three aspects of image enhancement preprocessing, attention mechanism, and GELU activation function in detail. The exposition includes the principle of the module and how the data operates in the module.

2.1 Image Enhancement Preprocessing

The LF subviews obtained by decoding the images captured by the LF camera have different luminance ranges from the edges to the center, and the edge

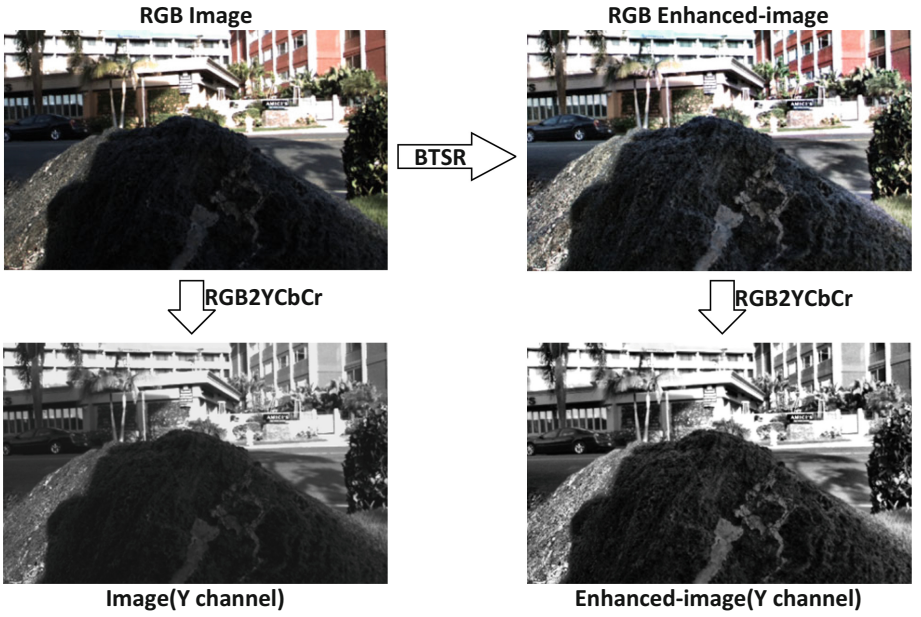

Fig. 3. The original image is compared with the enhanced image. The left side shows the RGB and Y channels of the original image, and the right side shows the RGB and Y channels of the enhanced image.

subviews will be dark. If the decoded image is directly used as the input, it will affect the reconstruction result of the network. This is because in the darker images, the texture information will be suppressed and it is difficult for the network to extract useful features from them. In particular, the high frequency parts of the image will become blurred due to the lower luminance. The boundaries between different objects become blurred, resulting in multiple objects being blended together. This causes distortion and ghosting in scene reconstruction in the presence of occlusion and complex textures. To avoid these problems, this paper uses BTSR to enhance the image, increasing the contrast between objects and making textures more visible.

In a single image, the correlation between pixels is inversely proportional to their distance in space. Therefore, the image $I(x, y)$ is first partitioned into k small blocks of $N \times N$ before image enhancement to obtain the set of local regions $R = \{R_1, R_2, ..., R_k\}$, where k is the number of local regions. Histogram equalization enhancement is applied to each block to obtain the enhanced local regions. The expression of enhancement as,

$$E_i(x, y) = T(R_i(x, y)), \tag{1}$$

where T is the mapping function that represents the mapping of the pixel value at position (x, y) in the local region R_i to the new pixel value $E_i(x, y)$. In order to avoid excessive increase in contrast, it is necessary to apply a contrast limit to each local area. The formula for the contrast limit is as follows:

$$CE_i(x, y) = \begin{cases} E_i(x, y), \text{if} \sigma_{\text{i}} \leq \text{H} \\ \frac{E_i(x,y)H}{\sigma_{\text{i}}}, \text{if} \sigma_{\text{i}} > \text{H} \end{cases}, \tag{2}$$

where σ_i is the standard deviation of the pixel values in the local region E_i and H is the specified threshold value. Finally this paper recombines the contrast-limited enhanced local region CE_i into the final enhanced image as,

$$F(x, y) = \begin{cases} CE_i(x, y), \text{if(x, y)} \in \text{R}_{\text{i}} \\ I(x, y), otherwise \end{cases}. \tag{3}$$

To simplify the training, this paper convert the enhanced RGB image into the format of YCbCr. Y denotes the intensity and brightness of the image, while Cb and Cr denote the blue chromaticity and red chromaticity of the image, respectively. The Y channel of the image has all the texture information of the original image, and this paper only need to use the Y channel image as the input during training. The effect of image enhancement is shown in Fig. 3, with the unenhanced image on the left and the enhanced image on the right. It can be clearly seen that the texture of the enhanced image is more obvious, both RGB and Y channel images.

2.2 HDAM

The core of the attention mechanism is resource allocation, which adjusts the allocation of resources according to the importance of the target in order to

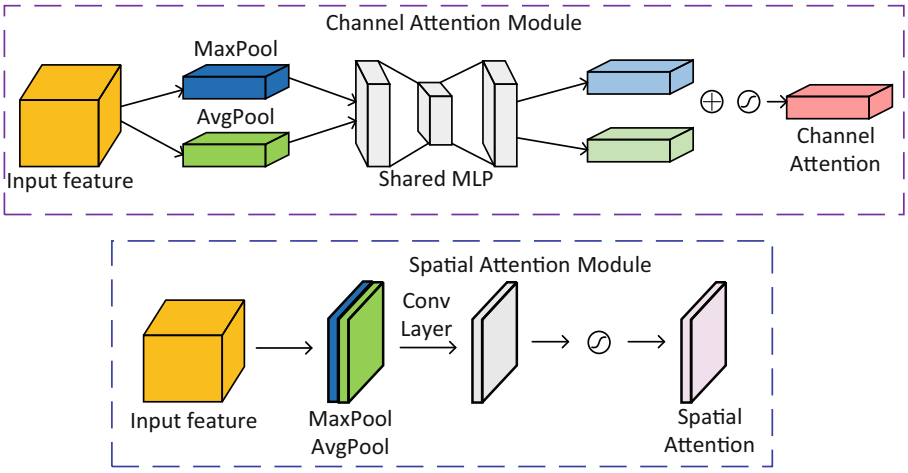

Fig. 4. Schematic diagram of each attention submodule. As shown in the figure, the channel attention utilizes the maximum pooled output and the average pooled output of the shared network. Spatial attention utilizes two similar outputs pooled along the channel axis and forwards them to the convolutional layer.

focus more on the important objects. In CNN, the attention mechanism adjusts the allocation of weight parameters. By allocating more weight parameters to the objects of attention, the representation of these objects is enhanced during feature extraction. Introducing the attention mechanism into the viewpoint synthesis task can improve the representation ability of the model and reduce the interference of irrelevant targets. Enhancing the reconstruction effect on the objects of attention and consequently improving the overall visual effect.

This paper introduces HDAM, an attention mechanism for CNN performance enhancement, into the network. It improves the expressive and perceptual capabilities of the model by introducing channel attention and spatial attention at different levels of the CNN.

As shown in the upper part of Fig. 4, channel attention is used to weight the feature maps of each channel to enhance the representation of important features. It converts each channel's feature map into a scalar by a global averaging pooling operation, and then learns it through two fully connected layers. Finally, the weighting factor is restricted between 0 and 1 using a Sigmoid function. In this way, the feature maps of each channel are multiplied by an attention weight to highlight the important features.

As shown in the lower part of Fig. 4, spatial attention is used to weight the different spatial locations of the feature maps to enhance the representation of important regions. It obtains two feature maps by performing maximum pooling and average pooling operations on each channel. Then they are connected and learned by a convolutional layer. Finally the weighting factor is restricted between 0 and 1 using a Sigmoid function. In this way, the features at each

spatial location are multiplied by an attention weight to highlight the important regions.

HDAM invokes channel attention and spatial attention sequentially along the output of the 4D convolutional feature map. The network is instructed to assign different attention to each feature module. Note that we place HDAM before upsampling, which ensures that the network's receptive field is large enough. It also balances the weight ratio between the perceptual field and the number of channels to a certain extent. The adaptive modification of the feature modules by HDAM allows for better retention and processing of important scenes in the image during upsampling.

2.3 GELU

During the training process, some pixels may be predicted to have negative values. Since all pixel values in the image should be greater than or equal to 0, this paper need to deal with the negative values that appear in the prediction. Generally, we think that negative-valued pixels around 0 still contribute to the image and this paper should keep them. On the other hand, negative pixels far from 0 should be excluded because they do not contribute much. Therefore, this paper chooses to use GELU as the activation function of the network, and its expression as,

$$GELU(X) = 0.5 \cdot x \cdot (1 + \tanh(\sqrt{\frac{2}{\pi}}(x + 0.044715x^3))). \tag{4}$$

Starting from the characteristics of the function itself, it treats pixel values in the way we expect. the GELU function is derivable throughout the real number domain and has continuous, smooth properties. This makes it easier to optimize in training and can provide faster convergence. Compared to LeakyReLU, GELU has a larger gradient in the region close to zero. This property makes GELU more advantageous in mitigating the gradient disappearance problem and facilitating gradient propagation. At the same time, the nonlinear property of GELU enables it to introduce more nonlinear transformations, thus providing stronger feature representation.

3 Experimental Results

Comprehensive experiments were conducted on real-world [22] and synthetic [23] scenes to verify the effectiveness of the proposed method. The 9×9 dense LF was reconstructed with 3×3 subviewpoint maps on the reconstruction task. the proposed method was compared with two LF angle reconstruction methods (HDDRNet and M-HDDRNet) proposed by Meng *et al.* [21]. This paper used the average PSNR and SSIM of the reconstructed SAI with its corresponding ground truth as performance criteria. For the proposed and compared methods, the positions of the input sparse 3×3 SAIs are shown in the red cells in the grid diagram.

3.1 Real-World Scenes

Table 1 shows the quantitative comparison results of our algorithm with HDDRNet and M-HDDRNet on the real-world dataset. As can be seen from the table, our proposed LF perspective SR network performs better in terms of target quality. Our method shows significant improvement in both the average PSNR metrics and also obtains consistent results on the average SSIM for all tested LF datasets. This is mainly due to the attention module we introduced during the network training process to guide the network to focus the reconstruction on scene features useful for SR. By using a more targeted reconstruction approach, we are able to obtain better reconstruction quality, especially in some fine-texture regions.

Figure 5 shows the visual comparison results of three real-world scenes reconstructed by SAI using three different methods. Reflective_29 is occupied by some reflective surfaces, Occlusions_9 contains many occluded regions, and scene Cars_7 has complex textures and large chromatic aberrations. As can be observed in Fig. 5, our proposed method is able to achieve a better perceptual quality of SAI reconstruction than the other two methods. Since the quality is susceptible

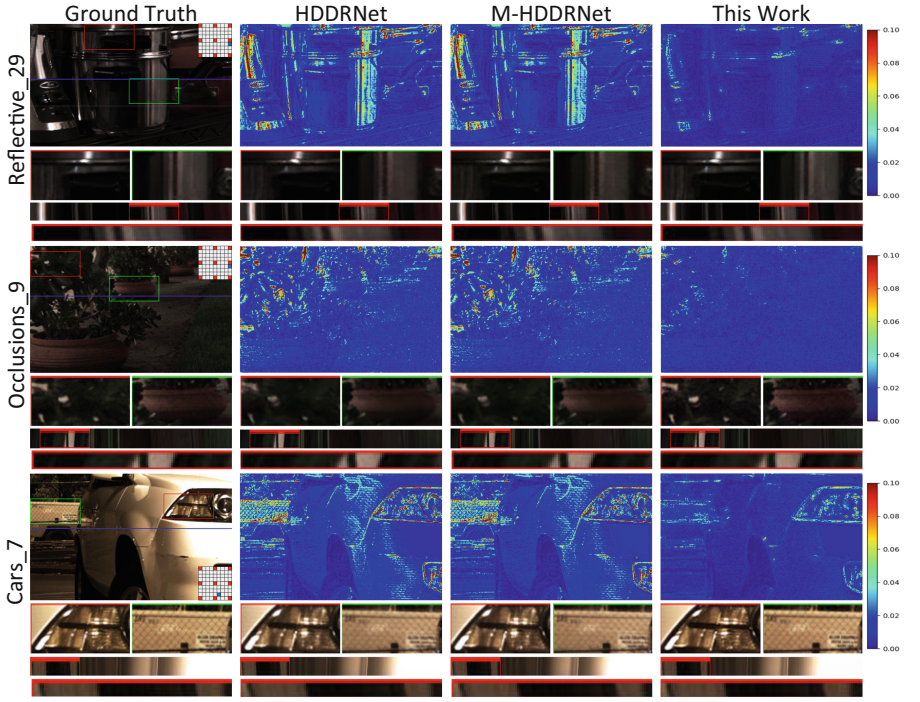

Fig. 5. A visual comparison of the three methods for the real-world perspective SR. The comparison shows the ground truth SAI, the error map of the Y-channel reconstructed SAI, the close-up version of the SAI part in the red and green boxes, and the EPI extracted at the blue line and its close-up. (Color figure online)

Table 1. Quantitative comparison of three algorithms for LF dataset Reflective, Occlusions and Cars reconstructions (PSNR/SSIM).

Algorithm	Reflective_29	Occlusions_9	Cars_7
HDDRNet	35.969/0.975	36.235/0.942	34.164/0.964
M-HDDRNet	36.021/0.976	36.276/0.943	34.219/0.964
This work	41.195/0.989	40.676/0.962	38.315/0.979

to parallax variations, it can be observed from the close-up images and error maps that HDDRNet and M-HDDRNet introduce blurring and ghosting artifacts. In contrast, our proposed method not only performs better in terms of reconstruction quality, but also has good results in processing and recovering details.

3.2 Synthetic Scenes

Table 2. Quantitative comparison of three algorithms for LF dataset Cotton, Dino and Tomb reconstruction (PSNR/SSIM).

Algorithm	Cotton_29	Dino_9	Tomb_7
HDDRNet	38.700/0.953	34.472/0.913	36.405/0.852
M-HDDRNet	38.822/0.955	34.615/0.916	36.468/0.855
This work	43.368/0.973	39.215/0.949	39.088/0.900

In order to verify the effectiveness of our proposed method in synthetic scenes, this paper conducted experiments and selected three synthetic scenes from the HCI dataset. These three scenes are Cotton, Dino and Tomb, which have different texture characteristics. The Cotton scene contains complex shadows and reflections, the Dino scene is rich in line textures, and the Tomb scene contains complex noise-like textures. We present the quantitative comparison results of the three methods in Table 2.

As can be observed from the table, our proposed method achieves higher PSNR than the other two methods in all scenes. Especially for Cotton scenes, our method achieves an average PSNR gain of up to 4.668 dB. This shows that our method performs well in LF synthetic scenes with smooth surfaces with self-obscuring shadows and reflections. For the Dino scene, this paper obtained consistent experimental results. This is due to the enhanced preprocessing we used during the network training process, which allows the network to perceive more texture details. In the Tomb scene, our method achieves a PSNR gain of 2.683 dB. This also indicates that our method has stronger resistance to interference for scenes with noisy textures.

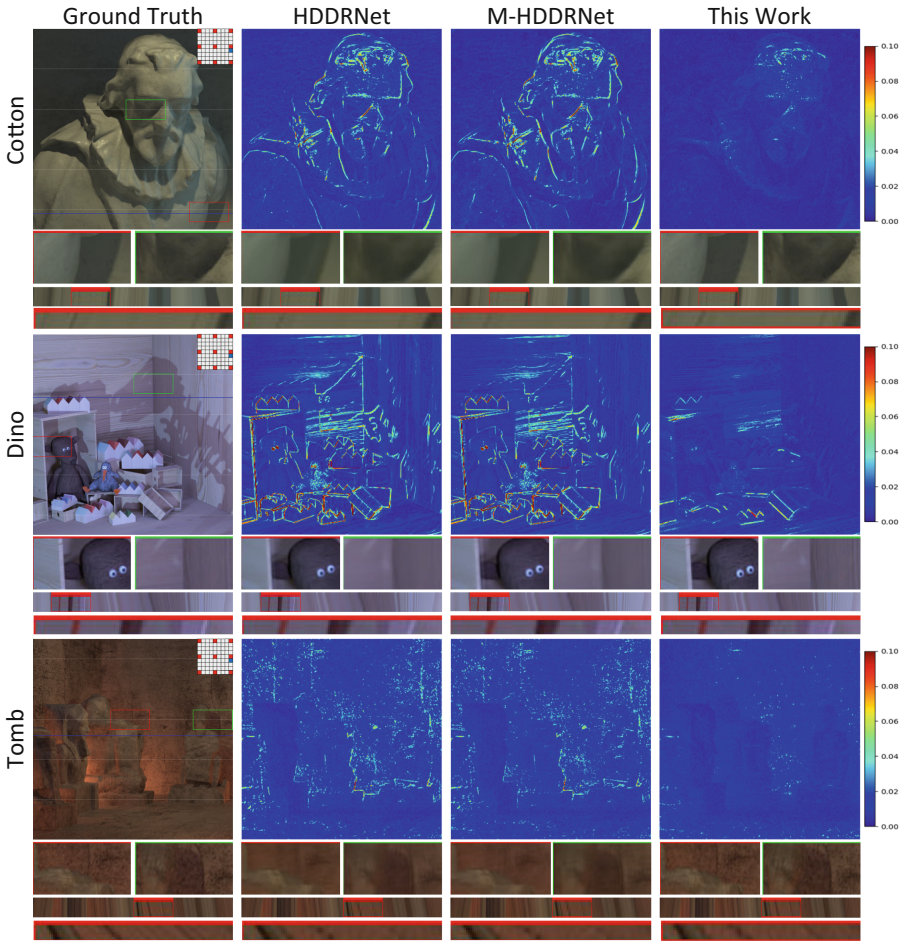

Fig. 6. A visual comparison of the three methods for the synthetic scene angle SR. The comparison shows the ground truth SAI, the error map of the Y-channel reconstructed SAI, the close-up version of the SAI part in the red and green boxes, and the EPI extracted at the blue line and its close-up. (Color figure online)

Figure 6 shows more visually the differences between the three methods in the reconstruction. Especially for the Tomb scene, this paper can clearly find that HDDRNet and M-HDDRNet do not reconstruct well at high frequencies and tend to lose high-frequency textures. And our proposed method achieves the recovery of many high-frequency detailed textures by exploring more texture information. Further observing the close-up images and error maps, we find that our method is also applicable to Dino and Cotton scenes, and especially performs well in the reconstruction of some textures and edge regions.

4 Conclusion

In this paper, we improve and replace some modules of 4D CNN in paper [21]. We enhance the image pre-processing in the pre-training stage, which enhances the image contrast and makes the texture clearer. The attention mechanism HDAM is introduced, and this module can be applied to 4D convolution to build the attention map along the direction of the original activation function is replaced to make the network converge faster. Finally, after an experimental comparison, our proposed method leads to a better performance of the network. In subsequent work we will consider lightweighting the model so that it can be applied to real-time tasks.

Acknowledgment. This work was supported in part by National Natural Science Foundation of China (No. 62067003), Culture and Art Science Planning Project of Jiangxi Province (No. YG2018042), Humanities and Social Science Project of Jiangxi Province (No.JC18224).

References

1. Shin, C., Jeon, H.-G., Yoon, Y., Kweon, I.S., Kim, S. J.: EPINET: a fully-convolutional neural network using epipolar geometry for depth from light field images. In Proceedings of IEEE Conference on Computer Vision Pattern Recognition, pp. 4748–4757 (2018)
2. Mitra, K., Veeraraghavan, A.: Light field denoising, light field superresolution and stereo camera based refocussing using a GMM light field patch prior. In: Proceedings of IEEE Conference on Computer Vision Pattern Recognition Workshops, pp. 22–28 (2012)
3. Yucer, K., Sorkine-Hornung, A., Wang, O., Sorkine-Hornung, O.: Efficient 3D object segmentation from densely sampled light fields with applications to 3D reconstruction. ACM Trans. Graph. **35**(3), 22:1–22:15 (2016)
4. Kim, C., Zimmer, H., Pritch, Y., Sorkine-Hornung, A., Gross, M.: Scene reconstruction from high spatio-angular resolution light fields. ACM Trans. Graph. **32**(4), 73:1–73:12 (2013)
5. Wang, T.-C., Efros, A.A., Ramamoorthi, R.: Occlusion-awaredepth estimation using light-field cameras. In: Proceedings of IEEE International Conference on Computer Vision, pp. 3487–3495 (2015)
6. Pearson, J., Brookes, M., Dragotti, P.L.: Plenoptic layer-based modeling for image based rendering. IEEE Trans. Image Process. **22**(9), 3405–3419 (2013)
7. Zhang, Z., Liu, Y., Dai, Q.: Light field from micro-baseline image pair. In: Proceedings of IEEE Conference on Computer Vision Pattern Recognition, pp. 3800–3809 (2015)
8. Zhang, F.-L., et al.: PlenoPatch: patch-based plenoptic image manipulation. IEEE Trans. Visualization Comput. Graph. **23**(5), 1561–1573 (2017). https://doi.org/10.1109/TVCG.2016.2532329
9. Chai, J.X., Tong, X., Chan, S.C., et al.: Plenoptic sampling. In: Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques, pp. 307–318 (2000)

10. Zhang, C., Chen, T.: Spectral analysis for sampling image-based rendering data. IEEE Trans. Circuits Syst. Video Technol. **13**(11), 1038–1050 (2003)
11. Do, M.N., Marchand-Maillet, D., Vetterli, M.: On the bandwidth of the plenoptic function. IEEE Trans. Image Process. **21**(2), 708–717 (2011)
12. Zhu, C.J., Yu, L.: Spectral analysis of image-based rendering data with scene geometry. Multimedia Syst. **23**, 627–644 (2017)
13. Vagharshakyan, S., Bregovic, R., Gotchev, A.: Light field reconstruction using shearlet transform. IEEE Trans. Pattern Anal. Mach. Intell. **40**(1), 133–147 (2018)
14. Chen, W., Zhu, C.: Spectral analysis of a surface occlusion model for image-based rendering sampling. Digital Signal Process. **130**, 103697 (2022)
15. Yoon, Y., Jeon, H.G., Yoo, D., Lee, J.Y., So Kweon, I.: Learning a deep convolutional network for light-field image superresolution. In: Proceedings of IEEE International Conference on Computer Vision Workshops, pp. 24–32 (2015)
16. Flynn, J., Neulander, I., Philbin, J., Snavely, N.: DeepStereo: learning to predict new views from the world's imagery. In: Proceedings of IEEE Conference on Computer Vision Pattern Recognition, pp. 5515–5524 (2016)
17. Wu, G., Liu, Y., Fang, L., Dai, Q., Chai, T.: Light field reconstruction using convolutional network on EPI and extended applications. IEEE Trans. Pattern Anal. Mach. Intell. **41**(7), 1681–1694 (2019)
18. Wu, G., Liu, Y., Dai, Q., Chai, T.: Learning sheared EPI structure for light field reconstruction. IEEE Trans. Image Process. **28**(7), 3261–3273 (2019)
19. Wang, Y., Liu, F., Wang, Z., Hou, G., Sun, Z., Tan, T.: End-to-end view synthesis for light field imaging with Pseudo 4DCNN. In: Proceedings of European Conference on Computer Vision, pp. 333–348 (2018)
20. Yeung, W.F.H., Hou, J., Chen, J., Chung, Y.Y., Chen, X.: Fast light field reconstruction with deep coarse-to-fine modeling of spatial-angular clues. In: Proceedings of European Conference on Computer Vision, pp. 137–152 (2018)
21. Meng, N., So, H.K.H., Sun, X., et al.: High-dimensional dense residual convolutional neural network for light field reconstruction. IEEE Trans. Pattern Anal. Mach. Intell. **43**(3), 873–886 (2019)
22. Raj, S., Lowney, M., Shah, R., Wetzstein, G.: Stanford lytro light field archive (2016). http://lightfields.stanford.edu/LF2016.html.
23. Honauer, K., Johannsen, O., Kondermann, D., Goldluecke, B.: A dataset and evaluation methodology for depth estimation on 4D light fields. In: Proceedings of Asian Conference on Computer Vision, pp. 19–34 (2016)