A Convolutional Neural Network Using Epipolar Geometry for Light Field Images Super Resolution

Yuxing Liu Fude Cao Yihao Geng Yuhan Zhang
Computer Science and Engineering
University of Michigan
Ann Arbor, United States
{fudecao, dylanliu, yhgeng, zyuhan}@umich.edu

Abstract

Super resolution for images are important image processing techniques for improving resolution of images and videos in computer vision field. While we are seeing an increased usage of light-field cameras, owing to the limitations of the sensor technology inside the light-field camera, the sub-aperture images acquired by the current light-field camera are generally of low resolution, which hinders the development of light-field imaging. To address this, we have implemented a neural network that processes sub-aperture images(SAIs) in four orientations to best exploit the disparity information inherently present in different viewing directions. We first use 4 dedicated convolutional blocks to learn disparity features individually. Then, a collect-and-distribute strategy merges them to boost the complementary information across views to improve the High-resolution SAIs. Furthermore, we propose a channel attention model integrated into the reconstruction block to suppress redundant information, thereby highlighting useful data and restoring high-frequency details. We demonstrate through comparative experiments over multiple datasets that our method can recover sharp edges and fine textures in images, outperforming most of the existing 2D image and bicubic interpolation algorithms regarding the quality of reconstruction. Besides, a minimal computational complexity network is presented, attaining state-of-the-art performance, specifically in dealing with disparity variations.

1. Introduction

Image Super Resolution (SR) is the recovery of a high-resolution image from a low-resolution image or sequence of images. High-resolution images contain rich information, and this rich information has a wide range of applications in image segmentation, image depth estimation, and image saliency detection. Light Field (LF) cameras record not only the intensity but also the direction of light by adding a microlens array between the main lens and the sensor. However, due to the limitation of the sensor technology inside the light field camera, the sub-aperture image resolution acquired by the current light field camera is generally low, and the problem of low spatial resolution hinders the development of light field imaging. High-resolution images are needed in practical applications, so it is necessary to improve the resolution of low-resolution light-field images and generate high-resolution light-field images.

EPInet [17], as a simple but effective network structure, has achieved good performance in depth estimation tasks. Although EPINET was originally designed for depth estimation, its fully convolutional neural network structure and efficient processing methods for light-field images also give it potential for applications in the field of super-resolution of light-field images. Unlike single-image super-resolution methods, EPINET is able to utilize the relationships between neighboring sub-aperture images in light-field images to better maintain the geometric structure and content consistency of the images. Therefore, we draw on the network method and architecture of EPINET and apply it to the light-field image super-resolution task to improve the spatial resolution of the sub-aperture images so as to obtain higher quality super-resolution light-field images. And for the light-field image super-resolution task, we design a reconfiguration module that uses residual connectivity and attention mechanisms to prevent the degradation of the network and improve the super-resolution performance.

2. Background

2.1. Technical background

In 1939, Gershun [5] formally defined the concept of the light field, laying the groundwork for the systematic study of spatial radiance distribution. Later, in 1991, Adelson and Bergen [1] proposed the plenoptic function, significantly enriching the theoretical framework of the light field. This function is represented by a seven-dimensional variable including position, angle, wavelength, and time, given by $LF = LF(x, y, z, \theta, \phi, \gamma, t)$. Wherein θ represents the intensity of light, ϕ represents direction, γ represents wavelength, and t represents time. However, due to the high dimensionality of seven-dimensional data, recording and processing in practical applications are challenging. As a result, researchers have simplified the seven-dimensional plenoptic function to enhance data processing efficiency and decrease the cost of recording. In 2006, Levoy et al. [10] proposed using a dual-plane model to parameterize the light field, allowing the four-dimensional parameters to be represented as $LF(x, y, s, t)$, where (x, y) denotes the coordinates on the spatial plane, indicating the imaging pixel location for a point within a scene, and (s, t) represents the coordinates on the angular plane, corresponding to the position from a particular viewpoint.

Since it is difficult to visualize the spatial structure of the four-dimensional data, in order to visualize the light field intuitively, it can be realized by fixing certain dimensions. Three manifestations of light field visualization will be introduced in detail: Array in Sub-Aperture Image(SAI), Macro-Pixel Image(MPI), and Polar Plane Image(PPI)1. (1) Sub-aperture images visualize the scene imaging at different viewing angles, arranged as an image array. By fixing the angular coordinates(s^*, t^*), the sub-aperture images $LF(x, y, s^*, t^*)$under a certain viewing angle can be obtained. The 4D light field can be represented as a sheet of $s \times t$ sub-aperture images and the resolution of each sub-aperture image is $x \times y$.

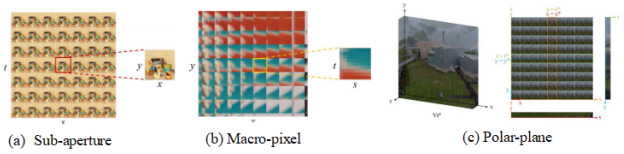

Figure 1. Three manifestations of light field visualization

(2) Macro-pixel image is another representation of light field visualization, which is known from the four-dimensional parameterization of light field, fixing the spatial coordinates(x^*, y^*), we can get the macro-pixel image $LF(x^*, y^*, s, t)$ of the same scene viewpoint imaged under different viewpoints. It represents the pixel values under the images of all sub-viewpoints in the scene, which contains all the rays emitted by the object, and the macro-pixel image contains rich angle information.

(3) Unlike subaperture images and microlens images, polar plane images contain both angular and spatial information. The EPI can be obtained by fixing one angular coordinate and one spatial coordinate in the 4-dimensional coordinates of the light field.

2.2. Related work

Mitra et al. [13] introduced a method for light field reconstruction combining the Linear Minimum Mean Square Error (LMMSE) estimator and Gaussian Mixture Model (GMM). Wanner et al. [24] proposed a variational method for parallax estimation, spatial and angular supersegmentation. Farrugia et al. [4] proposed the use of a linear subspace projection method for spatial over-segmentation of the light field. Rossi et al. [15] proposed a graph-based method that avoids the requirement for accurate parallax estimation. However, these methods take a long time to generate the final super-resolution light-field images and the quality of the reconstructed images is still limited compared to current deep learning-based methods.

Deep learning has achieved great success in the field of single-image super-resolution in recent years, and inspired by these works, recent super-resolution methods for light-field images employ deep convolutional networks to improve their performance. Yoon et al. [27] first used a convolutional neural network (CNN) to process light field images with the proposed LFCNN, which first improves the spatial resolution of neighboring sub-aperture image pairs via an SRCNN [3], and then synthesizes a new view between neighboring sub-aperture image pairs via a structurally similar SRCNN [3]. Some scholars applied single-image super-resolution methods-VDSR [8], EDSR [12], and RCAN [30]-to light-field super-resolution to improve the spatial resolution of individual subaperture images, but the use of single-image super-resolution ignores the relationship between neighboring subaperture images, resulting in poor image quality. Yuan et al. [28] utilized the characteristics of the extreme planar image (EPI) to propose the LF-DCNN. This network is divided into two parts: first, it spatially super-resolves each sub-aperture image using the EDSR [12] network; then, it extracts the EPI to input into the EPI enhancement network. The latter consists of several dense residual blocks and ultimately converts the reconstructed EPI into a high-resolution sub-aperture image. The 1st part of both LFCNN and LF-DCNN simply enhances the spatial information Wang et al. [20] introduced bi-directional recurrent convolutional network (BRCNN) to super-resolution of light field images and proposed LFNet, in which they built two bi-directional recurrent convolu-

tional sub-networks with the same structure to iteratively learn spatial correlation for neighboring views of one column and one row of sub-aperture images, respectively, and then fused the outputs of the two sub-networks to generate a high-resolution light field through stack generalization technique. Yeung et al. [26] proposed LFSSR, which utilizes alternating spatial and angular convolution to improve the super-resolution of all views. Although deep learning based methods have recently been proposed to achieve state-of-the-art performance, these methods obtain accuracy improvements at the expense of parametric counts.

3. Methodology

3.1. Library Used

The following libraries are utilized in this project:

Pytorch, Numpy, Scipy, Matplotlib, tqdm, time, argparse, random, h5py, skimage, os

3.2. Function used

The sigmoid activation function is widely used in neural networks. It can be expressed as:

$$\sigma(x) = \frac{1}{1 + e^{-x}} \tag{1}$$

where x is the input to the function, e is the base of the natural logarithm, and $\sigma(x)$ is the output between 0 and 1. The Leaky Rectified Linear Unit, known as the LeakyReLU, activation function in a neural network can be expressed as:

$$LeakyReLU(x) = \begin{cases} 0.1x & \text{if } x < 0 \\ x & \text{if } x \geq 0 \end{cases}$$

The Batch Normalization for a 2D input in a neural network can be expressed as:

$$y = \left(\frac{x - \mathrm{E}[x]}{\sqrt{\sigma_B^2 + \epsilon}} \right) \gamma + \beta$$

where:

- x is the input to a layer,
- μ_B is the mean of the batch,
- σ_B^2 is the variance of the batch,
- ϵ is a small constant added for numerical stability (to avoid division by zero),
- γ and β are parameters to be learned,
- y is the normalized output.

In 4D space, the point cloud $L(x, y, s, t)$ in $R^{S \times T \times X \times Y}$, where (x, y) represents the spatial coordinates and (s, t) represents the temporal frequency coordinates. The network designed in this article takes a low-resolution sub-aperture image array $L^{LR}(x, y, s, t) \in R^{S \times T \times X \times Y}$ as input, which, after a learnable function f, produces a high-resolution sub-aperture image array as $L^{HR}(\overline{x}, \overline{y}, s, t) \in R^{S \times T \times \alpha X \times \alpha Y}$, where α represents the up-sampling factor.

3.3. Network Architecture Details

Compared with 2D images, light-field images have richer four-dimensional information, and the core problem of light-field image super-segmentation is to make full use of the angular and spatial information and recover the texture details to improve the image quality. The overall structure of the proposed network- although ideas may inspired by the paper cited above, the codes were completely implemented by ourselves- is shown in Figure2. where the network takes the sub-aperture image as input and acquires image stacks in four directions: horizontal, vertical, left diagonal, and right diagonal, respectively, and the image stacks in each direction are respectively activated by one convolution block - "convolution-activation-convolution-normalization-activation " to acquire features in their respective directions. Subsequently, the feature channels in the 4 directions are spliced together, and then the fusion features are obtained by learning complementary information through 8 cascaded convolutional blocks. Finally, the fused features are passed through a reconstruction module and an upsampling module to obtain a high-resolution sub-aperture image array.

In the convolutional block, we use two different sizes of convolutional kernels: 1×1 and 3×3. In extracting the features in the four directions of the image stack, the 1×1 convolution is first used to perform the initial extraction of the features and to boost the number of channels, followed by the 3×3 convolution for further extraction of the features to ensure that richer information is captured. In the fusion stage, we use 1×1 convolution to compress the number of channels to reduce the number of parameters, while 3×3 convolution is used to expand the number of channels to better mimic the "squeeze" and "excite" operations in the attention mechanism. This design aims at capturing the inter-channel information. This design aims to capture the relationship between channels, thus reducing the number of parameters while maintaining the super-resolution performance and improving the performance and efficiency of the network. With this strategy, we are able to better optimize the network structure to improve the performance and performance of the super-resolution task for light-field images.

The reconstruction module cascades two sets of residual blocks (Resblocks) [6] and channel attention modules [16], with the network structures of RB and CA illustrated in Figures Figure 4. and Figure 3., respectively. The Resblock is

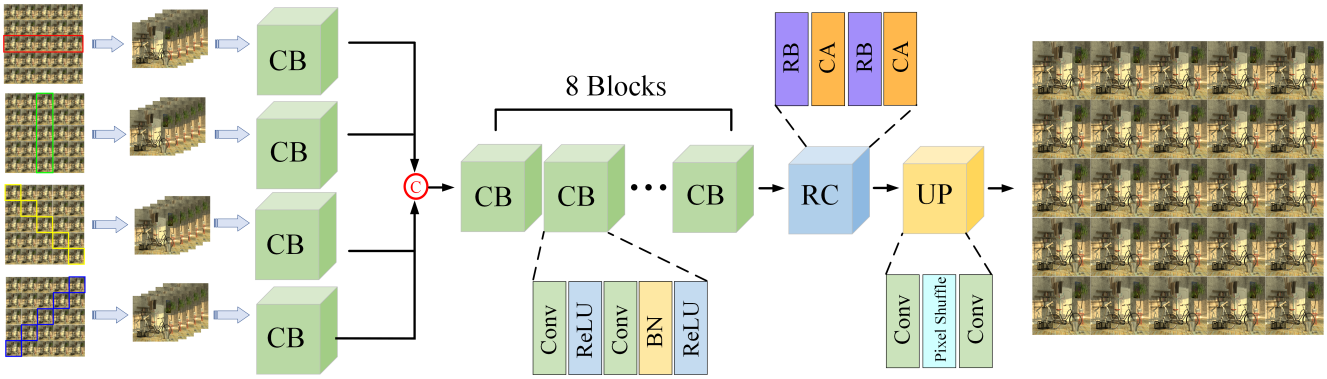

Figure 2. Architecture of the our model

composed of two 3×3 convolutions followed by a Leaky ReLU activation function. By employing a residual connection approach, the block effectively merges primary and high-level features, thus addressing the degradation problem within the network. The channel attention module utilizes parallel adaptive average pooling and max pooling to condense spatial information. Subsequently, a 1×1 convolution compresses the channel dimension, which, after the application of a ReLU activation function, is expanded again by another 1×1 convolution. To capture inter-channel dependencies, two 1×1 convolutions are employed to upscale and downscale the results of both pooling methods. Lastly, channel weights are generated by a Sigmoid activation function and then applied to the input feature maps to modulate the channel-wise features.

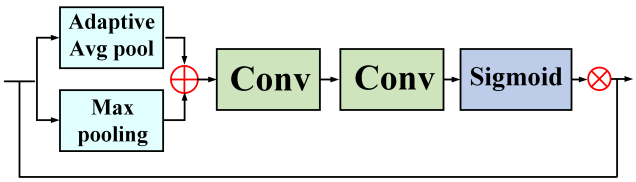

Figure 3. Channel Attention Block

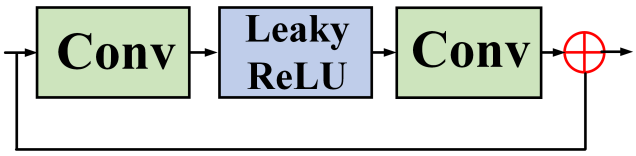

Figure 4. Residual Block

In the model's up-sampling module, the feature maps after the reconstruction blocks are projected onto a smaller dimension space by a 1×1 convolution, extending the feature map dimensions to $\alpha^2 C$. Following this, a Pixel Shuffle operation is employed to upscale the feature map by a factor of $\alpha X \times \alpha Y$. Ultimately, another 1×1 convolution is utilized to adjust the number of channels, ensuring the output feature map aligns with the required dimensions.

4. Experiments

4.1. Dataset and training details

The experiment uses five publicly available datasets: EPFL [14], HCI New [7], HCI Old [25], INRIA [9], and STFgantry [18], to evaluate the performance of the algorithm. These datasets contain 5 to 9 scenes of various views, each providing a low-resolution sub-aperture image array and its corresponding high-resolution version. The images are pre-processed to a uniform size of 64x64 pixels. The network is implemented using Pytorch and optimized with the Adam optimizer. During training, the batch size is set to 8, and the learning rate is initially set to 2×10^{-4}, with the learning rate halved every 50 epochs until it reaches a minimum value.

Datasets	Training	Test
EPFL [14]	70	10
HCInew [7]	20	4
HCIold [25]	10	2
INRIA [9]	35	5
STFgantry [18]	9	2
Total	144	23

Table 1. Number of Datasets used in training and testing phase

4.2. Quantitative analysis

In the experiments, the Bicubic is used as a benchmark to compare this paper's method with the traditional single-image super-resolution algorithms VDSR [8], EDSR [12], and deep learning-based lightfield super-resolution algo-

rithms LFBM5D [2], GB [15], RCAN [30], ResLF [29]. The results of quantitative comparison between this paper's algorithm and other algorithms under different super-segmentation tasks are shown in Table 2, using peak signal-to-noise ratio (PSNR) [23] and structural similarity (SSIM) [22] as evaluation metrics. The metric scores for other algorithms are derived from the study conducted by Yingqian Wang [21].

As can be seen from the table, under the 2-fold superscoring task, the PSNR [23] and SSIM [22] values of the proposed algorithm in this paper exceed the baseline model (Bicubic) by 3.43dB and 0.0256, respectively, and at the same time, it exceeds both the traditional algorithms and some of the deep learning-based algorithms. Under the 4-fold superscoring task, the PSNR [23] and SSIM [22] values of the proposed algorithm exceed the baseline model by 3.35dB and 0.0271, respectively, and are maximized on all 5 datasets, which verifies the effectiveness of the proposed algorithm.

4.3. Qualitative inorganic analysis

Figure 5. shows the comparison of the visual effect of different algorithms on the "Origami" reconstruction of the scene under the 2-fold super-scoring task. As can be seen from the figure, in the texture-intensive part, the algorithm of this paper recovers a higher quality image, which can recover the "bar-like" structure in the scene. The reason is that the VDSR [8] and EDSR [12] algorithms over-segment a single subaperture image, ignoring the angle information. Figure 6. shows the comparison of the common visual effects of different algorithms on the scene "Bedroom" under the 4-fold super-segmentation task. As can be seen from the figure, in the area framed by the red box, the image quality recovered by this paper's algorithm is significantly better than that of Bicubic and other deep learning-based algorithms, and the reconstructed image in this paper is closer to GroudTruth, which further verifies the superiority of the algorithms proposed in this paper.

4.4. Complexity analysis

After verifying the performance of the proposed algorithm for over-scoring, the complexity is further analyzed.The comparison of Parameters and FLOPs of different algorithms under 4-fold over-scoring task is shown in Table 3.

Table 3. Comparison of Parameters and FLOPs of different algorithms under 4-fold super-resolution task

Method	Scale	Parameters	FLOPs (G)	PSNR/SSIM
EDSR [12]	4×	38.89	40.66×10^{25}	30.20/0.9121
RCAN [30]	4×	15.36	15.65×10^{25}	30.27/0.9133
ResLF [29]	4×	6.79	39.70	30.43/0.9223
Ours	4×	0.88	2.81	30.93/0.9303

From the table, it can be seen that this paper's algorithm has the smallest values of Parameters and FLOPs, but the highest PSNR [23] and SSIM [22] values are obtained.This paper's algorithm obtains the highest accuracy with the minimum complexity, which further validates the effectiveness of the proposed algorithm.

5. Conclusion

In this paper, we propose A Convolutional Neural Network Using Epipolar Geometry for Light Field Images Super Resolution, which makes full use of the complementary information on different angles to enhance the reconstruction quality. The network takes subaperture images in four directions, namely horizontal, vertical, left diagonal and right diagonal, as input, and learns the parallax information of the subaperture image arrays in the four directions through four independent convolutional blocks. Subsequently, the information in the four directions is combined to learn the complementary information in the four directions through convolutional blocks. To solve the problem of information redundancy, a channel attention module is added to suppress redundant information, highlight useful information, and recover high-frequency details. Experimental results on multiple light-field datasets show that the proposed algorithm can effectively recover the high-frequency details of the image, make the edges clearer, and get better reconstruction results than most of the existing 2D images and bicubic interpolation algorithms.

5.1. Future Work

In this paper, we have only utilized some of the views in the sub-aperture image array and have not fully utilized all of the views in the sub-aperture image array, resulting in insufficient information exploration. In order to improve the super-resolution performance, we will fully utilize the 4D properties of the light-field images in the next work by considering the relationships between the sub-aperture images as well as the relationships within the sub-aperture images. This means that we will explore the properties of light-field images more deeply to improve the performance of the super-resolution algorithm by integrating the spatial and angular information of the sub-aperture images. At

Table 2. Comparison of PSNR [23]/SSIM [22] values of different algorithms under 2x and 4x tasks

		Dataset					
Method	Scale	**EPFL** [14]	**HCInew** [7]	**HCIold** [25]	**INRIA** [9]	**STFgantry** [18]	Average
Bicubic	2×	29.74/0.9376	31.89/0.9356	37.69/0.9785	31.33/0.9577	31.06/0.9498	32.34/0.9518
LFBM5D [2]	2×	31.15/0.9545	33.72/0.9548	39.62/0.9854	32.85/0.9695	33.55/0.9718	34.18/0.9672
GB [15]	2×	31.22/0.9591	35.25/0.9692	40.21/0.9879	32.76/0.9724	35.44/0.9835	34.98/0.9744
VDSR [8]	2×	32.50/0.9599	34.37/0.9563	40.61/0.9867	34.43/0.9742	35.54/0.9790	35.49/0.9712
EDSR [12]	2×	33.09/0.9631	34.83/0.9594	41.01/0.9875	34.97/0.9765	36.29/0.9819	36.04/0.9728
Ours	2×	32.96/0.9696	35.65/0.9673	40.30/0.9878	34.42/0.9800	35.54/0.9821	35.77/0.9774
Bicubic	4×	25.14/0.8311	27.61/0.8507	32.42/0.9335	26.82/0.8860	25.93/0.8431	27.58/0.8661
LFBM5D [2]	4×	26.61/0.8689	29.13/0.8823	34.23/0.9510	28.49/0.9137	28.30/0.9002	29.35/0.9032
GB [15]	4×	26.02/0.8628	28.92/0.8842	33.74/0.9497	27.73/0.9085	28.11/0.9014	28.90/0.9013
VDSR [8]	4×	27.25/0.8782	29.31/0.8828	34.81/0.9518	29.19/0.9208	28.51/0.9012	29.81/0.9070
EDSR [12]	4×	27.84/0.8858	29.60/0.8874	35.18/0.9538	29.66/0.9259	28.70/0.9075	30.20/0.9121
RCAN [30]	4×	27.88/0.8863	29.63/0.8880	35.20/0.9540	29.76/0.9273	28.90/0.9110	30.27/0.9133
ResLF [29]	4×	27.46/0.8899	29.92/0.9011	36.12/0.9651	29.64/0.9339	28.99/0.9214	30.43/0.9223
Ours	4×	28.17/0.9052	30.38/0.9070	36.44/0.9674	30.32/0.9448	29.34/0.9269	30.93/0.9303

Figure 5. Origami

the same time, we will also look for and try more existing super-resolution networks for light-field images, such as LF-DFnet [21], EPIT [11] and DPT [19], and try to reproduce their codes as well as make improvements to achieve a higher level of super-resolution performance. For example, adjusting the network architecture, optimizing the hyperparameters, and improving the loss function. Through continuous exploration and experimentation, we will aim to improve the effectiveness of the super-resolution algorithms. It is suggested that in future research, in addition to considering the properties of the light field itself, we can also try to explore other mechanisms, single-image super-resolution algorithms and multimodal algorithms to improve the super-resolution performance. For example, the use of attention mechanisms in deep learning can be considered to improve the model's focus on key information in the image, or to combine traditional image processing methods to deal with specific types of light field images. These approaches may bring new breakthroughs and advances in the field of light-field image super-resolution.

References

[1] Edward H. Adelson and James R. Bergen. *The plenoptic function and the elements of early vision*. Cambridge, MA, USA, 1991. 2

[2] Marcus Alain and Aljoscha Smolic. Light field super-resolution via lfbm5d sparse coding. In *2018 25th IEEE International Conference on Image Processing (ICIP)*, pages 2501–2505. IEEE, 2018. 5, 6

[3] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Image super-resolution using deep convolutional networks, 2015. 2

[4] R A Farrugia, C Galea, and C Guillemot. Super resolution of light field images using linear subspace projection of patch-

Figure 6. Bedroom

volumes. *IEEE Journal of Selected Topics in Signal Processing*, 11(7):1058–1071, 2017. 2

[5] A Gershun. The light field. *Journal of Mathematics and Physics*, 18(1-4):51–151, 1939. 2

[6] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. 3

[7] K Honauer, O Johannsen, D Kondermann, et al. A dataset and evaluation methodology for depth estimation on 4d light fields. In *Computer Vision–ACCV 2016: 13th Asian Conference on Computer Vision*, pages 19–34. Springer International Publishing, 2017. 4, 6

[8] J. Kim, J. K. Lee, and K. M. Lee. Accurate image super-resolution using very deep convolutional networks. In *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pages 1646–1654, 2016. 2, 4, 5, 6

[9] M Le Pendu, X Jiang, and C Guillemot. Light field inpainting propagation via low rank matrix completion. *IEEE Transactions on Image Processing*, 27(4):1981–1993, 2018. 4, 6

[10] Marc Levoy, Ren Ng, Andrew Adams, Matthew Footer, and Mark Horowitz. Light field microscopy. *ACM Transactions on Graphics*, 25(3), 2006. 2

[11] Zhengyu Liang, Yingqian Wang, Longguang Wang, Jungang Yang, Shilin Zhou, and Yulan Guo. Learning non-local spatial-angular correlation for light field image super-resolution. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 12376–12386, 2023. 6

[12] B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee. Enhanced deep residual networks for single image super-resolution. In *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, pages 136–144, 2017. 2, 4, 5, 6

[13] Kaushik Mitra and Ashok Veeraraghavan. Light field denoising, light field superresolution and stereo camera based refocussing using a gmm light field patch prior. In *2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, pages 22–28. IEEE, 2012. 2

[14] M Rerabek and T Ebrahimi. New light field image dataset. In *8th International Conference on Quality of Multimedia Experience (QoMEX)*, 2016. 4, 6

[15] Mattia Rossi and Pascal Frossard. Geometry-consistent light field super-resolution via graph-based regularization. *IEEE Transactions on Image Processing*, 27(9):4207–4218, 2018. 2, 5, 6

[16] Woo S, Park J, Lee J Y, et al. Cbam: Convolutional block attention module. *Proceedings of the European conference on computer vision (ECCV)*, 3(9), 2018. 3

[17] Shin, Changha, et al. Epinet: A fully-convolutional neural network using epipolar geometry for depth from light field images. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018. 1

[18] V Vaish and A Adams. The (new) stanford light field archive. Tech. Rep. vol. 6, no. 7, Comput. Graph. Lab., Stanford Univ., 2008. Available: http://lightfield.stanford.edu/. 4, 6

[19] Shunzhou Wang, Tianfei Zhou, Yao Lu, and Huijun Di. Detail-preserving transformer for light field image super-resolution. In *Proceedings of the AAAI conference on artificial intelligence*, volume 36, pages 2522–2530, 2022. 6

[20] Y Wang, F Liu, K Zhang, et al. Lfnet: A novel bidirectional recurrent convolutional neural network for light-field image super-resolution. *IEEE Transactions on Image Processing*, 27(9):4274–4286, 2018. 2

[21] Yingqian Wang, Jungang Yang, et al. Light field image super-resolution using deformable convolution. *IEEE Transactions on Image Processing*, pages 1057–1071, 2020. 5, 6

[22] Sheikh H R Wang Z, Bovik A C et al. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 14(4):600–612, 2004. 5, 6

[23] Sheikh H R Wang Z, Bovik A C et al. Scope of validity of psnr in image/video quality assessment. *Electronics letters*, 44(13):800–801, 2008. 5, 6

[24] Sven Wanner and Bastian Goldluecke. Variational light field analysis for disparity estimation and super-resolution. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(3):606–619, 2014. 2

[25] S Wanner, S Meister, and B Goldluecke. Datasets and benchmarks for densely sampled 4d light fields. In *VMV*, volume 13, pages 225–226, 2013. 4, 6

[26] H W F Yeung, J Hou, X Chen, et al. Light field spatial super-resolution using deep efficient spatial-angular separable convolution. *IEEE Transactions on Image Processing*, 28(5):2319–2330, 2018. 3

[27] Yongjin Yoon, Heung Gil Jeon, Donggeun Yoo, et al. Learning a deep convolutional network for light-field image super-resolution. In *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pages 24–32, 2015. 2

[28] Yucheng Yuan, Zhiwei Cao, and Liang Su. Light-field image superresolution using a combined deep cnn based on epi. *IEEE Signal Processing Letters*, 25(9):1359–1363, 2018. 2

[29] Shuo Zhang, Youfang Lin, and Hao Sheng. Residual networks for light field image super-resolution. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 11046–11055, 2019. 5, 6

[30] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu. Image super-resolution using very deep residual channel attention networks. In *Proc. Eur. Conf. Comput. Vis. (ECCV)*, pages 286–301, 2018. 2, 5, 6

6. Appendix

6.1. Code

Link to our code and results: download
(Access restricted to Umich accounts)

6.2. Details of First *FeaExtract* Block

Layer Type	**I**	**O**	**K**	**S**	**P**	**B**
Convolutional Layer	5	70	1	1	0	F
ReLU Activation	-					
Convolutional Layer	70	70	3	1	1	F
Batch Normalization	-					
ReLU Activation	-					

Table 4. Feature Extraction Stage: First *FeaExtract* Block
Legend: I = Input channels, O = Output channels, K = Kernel size, S = Stride, P = Padding, B = Bias (T/F)

6.3. Details of Second *FeaExtract* Block

Layer Type	**I**	**O**	**K**	**S**	**P**	**B**
Convolutional Layer	280	140	1	1	0	F
ReLU Activation	-					
Convolutional Layer	140	280	3	1	1	F
Batch Normalization	-					
ReLU Activation	-					

Table 5. Feature Extraction Stage: Second *FeaExtract* Block
Legend: I = Input channels, O = Output channels, K = Kernel size, S = Stride, P = Padding, B = Bias (T/F)

6.4. Details of *Residual* Block

Legend: I = Input channels, O = Output channels, K = Kernel size, S = Stride, P = Padding, B = Bias (T/F)

6.5. Details of *Attention Layer*

Legend: I = Input channels, O = Output channels, K = Kernel size, S = Stride, P = Padding, g = a grouping factor, typically set to 16

Layer Type	**I**	**O**	**K**	**S**	**P**	**B**
Convolutional Layer	70	70	3	1	1	F
LeakyReLU Activation	-					
Convolutional Layer	70	70	3	1	1	F

Table 6. Feature Blending Stage: *Residual* Block

Layer Type	**I**	**O**	**K**	**S**	**P**
Adaptive Avg Pooling	-	(1,1)	-	-	-
Adaptive Max Pooling	-	(1,1)	-	-	-
Add	-	-	-	-	-
Convolutional Layer	70	70/g	1	1	0
ReLU Activation	-				
Convolutional Layer	70/g	70	1	1	0
Sigmoid Activation	-				
Multiplication	-	-	-	-	-

Table 7. Attention Layer Structure

6.6. Details of *Upscale* Block

Layer Type	**I**	**O**	**K**	**S**	**P**	**B**
Convolutional Layer 1	70	70*f*f	1	1	0	F
Pixel Shuffle Layer	-					
Convolutional Layer 2	70	1	1	1	0	F

Table 8. Feature Blending Stage: *Upscale* Block

Legend: I = Input channels, O = Output channels, K = Kernel size, S = Stride, P = Padding, B = Bias (T/F)