Learning Epipolar-Spatial Relationship for Light Field Image Super-Resolution

Ahmed Salem, Hatem Ibrahem, Hyun-Soo Kang*
School of Information and Communication Engineering
Chungbuk National University, Cheongju, Korea
{ahmeddiefy, hatem, hskang}@cbnu.ac.kr

Abstract

Light field (LF) imaging has become increasingly popular in recent years for capturing and processing visual information. A significant challenge in LF processing is super-resolution (SR), which aims to enhance the resolution of low-resolution LF images. This article proposes a new LF image super-resolution (LFSR) approach that leverages the epipolar-spatial relationship within the LF. To train a deep neural network for LFSR, the proposed method involves extracting three types of information from the LF: spatial, horizontal epipolar, and vertical epipolar. Experimental results demonstrate the effectiveness of the proposed approach compared with state-of-the-art (SOTA) performance, as evidenced by quantitative metrics and visual quality. In addition, we conducted ablation studies to assess the effectiveness of each type of information and gain insights into the underlying mechanisms of the proposed method. Our approach achieved competitive results on the NTIRE 2023 Light Field Image Super-Resolution Challenge: our proposed model was ranked 10th on the test set and 6th on the validation set among 148 participants. Paper's code is available at: https://github.com/ahmeddiefy/EpiS_LFSR.

1. Introduction

Light field (LF) can record 3D geometry conveniently and efficiently (by recording the light intensity and directional information). As LF cameras have become so widespread, LF imaging has attracted the interest of researchers in both industry and academia. A few of the vision applications made possible by the wealth of data collected by LF camera photos include depth estimation [1, 2], salient object identification [3, 4], de-occlusion [5, 6], and others. However, due to the inherent angular-spatial tradeoff, the LF camera can either capture images with high spatial resolution but limited angular information or provide a large amount of angular information but with low spatial resolution [7]. Therefore, these limitations of the LF image sensor's resolution constrain the performance of the algorithms used in vision applications [8]. Several techniques have been introduced to improve the angular or spatial resolution of light field camera images to address this issue [1, 9]. This article concentrates explicitly on enhancing the spatial resolution of low-resolution LF images through LF image super-resolution (LFSR).

Several deep learning algorithms [10–17] based on different network architectures have recently been developed to improve LFSR using different LF datasets [18–22]. Convolutional neural networks (CNNs) and transformer-based networks conduct learning-based super-resolution via cross-view correlation. However, despite the improved LFSR performance, most of these algorithms do not adequately exploit the rich angular information, resulting in performance loss, particularly in complicated settings. Several approaches have attempted to extract epipolar, spatial, and angular information to improve image quality, but they still have limits.

Of the recently proposed methods to improve the quality of LFSR, Wang et al. [1] proposed a generic algorithm to analyze LF structure. They used it to achieve LFSR, LF reconstruction, and depth estimation. Liang *et al*. [23] proposed a transformer-based network including angular and spatial transformers to fully exploit the LF information and mitigate the effect of the small receptive field of CNN-based solutions. In their recent work, Liang *et al*. [24] proposed another transformer-based solution to increase the receptive field by learning the non-local spatial-angular correlation in LFs.

In this paper, we follow the pipeline of DistgSSR [1] and adopt the residual-in-residual structure. Precisely, we extract three types of information from the LF: spatial, horizontal epipolar, and vertical epipolar. Within each view, spatial information is extracted to exploit local context knowledge and long-range spatial relationships. In contrast, the epipolar information is extracted to learn the angular dependencies and understand the spatial-angular relation-

*Corresponding author

ship. Finally, we validate the performance of our proposed method through quantitative and qualitative (i.e., we use both synthetic and real-world LF datasets and depth estimation to evaluate the angular consistency) comparisons with the SOTA method, in addition to the achieved results on the NTIRE 2023 Light Field Image Super-Resolution Challenge.

The NTIRE 2023 Light Field Image Super-Resolution Challenge [25] seeks novel solutions for enhancing the spatial resolution of LF images. A total of 148 people have signed up for the challenge, and 11 groups have submitted results that surpass the PSNR scores of the baseline method LF-InterNet [15]. In addition, the newly proposed methods have established new SOTA standards in the field of LFSR. We make significant contributions in this work:

- We introduce our epipolar-spatial network, ranked 10th on the test set and 6th on the validation set.
- We validate the proposed approach through ablation studies and comparisons with the SOTA approaches.

2. Related Work

Light Field Image Super-Resolution (LFSR) creates high-resolution from low-resolution LF input images. One way to enhance the spatial resolution of the LF images is by using single-image super-resolution (SISR) techniques to view images independently. However, this approach fails to consider the correlation between different views, resulting in unsatisfactory results. Therefore, understanding the spatial-angular correlation among different views is important to achieve better LF spatial super-resolution.

Of the traditional methods, Wanner *et al*. [26] estimated the underlying disparity map using a variational technique that included LF angular and spatial information. To increase the accuracy of the disparity estimation, they additionally use total variation regularization. Mitra *et al*. [27] denoise, super-resolve, and refocus the pictures using the Gaussian mixture model (GMM) prior to the LF data. Farrugia *et al*. [28] partition low-resolution LF pictures into patches and stack them to produce patch volumes projected onto a high-dimensional subspace to increase the high-frequency information. Alain *et al*. [29] denoised the LF data using the LFBM5D technique and utilized the resulting denoised pictures as input for the sparse coding procedure. Sparse coding, on the other hand, is conducted individually on each view picture. Finally, Rossi *et al*. [30] create a graph out of LF data, with each node being an LF patch and the edges representing the similarity between the patches. Then, they utilized this graph to regularize the super-resolution process to maintain geometric uniformity among view pictures.

It has been demonstrated that deep learning-based algorithms perform better in LFSR. For example, Zhang *et al*. [10] employed a sub-aperture image alignment network (SAIN) to align the sub-aperture views before passing them through the ResNets. Yeung *et al*. [13] used a deep, efficient spatial-angular separable convolution (SASConv). The SASConv is designed to separately process the spatial and angular dimensions of the light field data. Jin *et al*. [12] proposed a method that also regularises structural consistency to ensure geometric consistency between the sub-aperture views. Wang *et al*. [15] enhanced the resolution of LF images by exploiting the interaction between the spatial and angular dimensions. Wang *et al*. [14] employed a network architecture that uses deformable convolution to model the geometric relationship between the sub-aperture views. Zhang *et al*. [11] utilized multiple epipolar geometries to extract information from neighbouring views for LFSR. Liu *et al*. [16] modelled the interaction between the sub-aperture views at both intra-view and inter-view levels. Of the recently proposed method to improve the quality of LFSR, Wang *et al*. [1] proposed a generic algorithm to analyze LF structure. They used it to achieve LFSR, LF reconstruction, and depth estimation. Liang *et al*. [23] proposed a transformer-based network including angular and spatial transformers to fully exploit the LF information to mitigate the effect of the small receptive field of CNN-based solutions. In their recent work, Liang *et al*. [24] proposed another transformer-based solution to increase the receptive field by learning the non-local spatial-angular correlation in LFs.

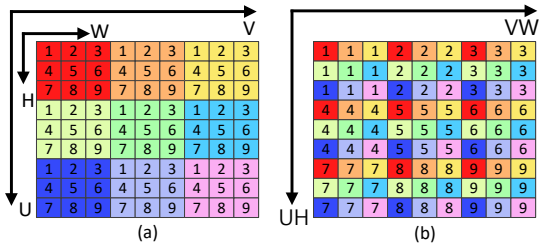

Figure 1. The relationship between SAI (a) and MacPI (b). Here, an LF of angular dimension $U{=}V{=}3$ and spatial dimension $H{=}W{=}3$. Different SAIs are painted with different colours, while different MacPI is denoted with different numbers.

3. Methodk

3.1. Problem Formulation

In this paper, we represent the LF as a 4D tensor (2D for the spatial dimension and 2D for the angular dimension), following the two-plane model [31]:

$$L(u, v, h, w) \in R^{U \times V \times H \times W} \tag{1}$$

Where (U, V) represent the 2D angular dimensions and (H, W) represent the 2D spatial dimensions. Following the SOTA methods [1, 23, 24], given a low-resolution (LR) LF: $L_{LR} \in R^{U\times V\times H\times W}$, we aim to upsample it to a high-resolution (HR): $L_{HR} \in R^{U\times V\times \alpha H\times \alpha W}$. Where the spatial resolution of the LR input is (H, W) and the spatial resolution of the HR output is (αH, αW), and α=2, 4 represents the upsampling or super-resolution factor.

The input LF is represented by a square array of Sub-Aperture Images (SAIs), but we analyze LFs using the Macro-Pixel Image (MacPI), with the mapping between Sub-Aperture Image (SAI) and MacPI shown in Figure 1.

3.2. LF Feature Extractors

We need to take the 4D structure of LF into consideration and design a network to model its non-local properties. Therefore, we extract the spatial information to exploit local context knowledge and long-range spatial relationships. In addition, the epipolar information is extracted to learn the angular dependencies and understand the spatial-angular relationship.

3.2.1 Spatial Feature Extractor

To exploit local context knowledge and long-range spatial relationships, DistgSSR utilizes a spatial feature extractor to process pixels within the same view and separate them from other views [1]. Following DistgSSR, a 3×3 convolution filter with a stride of 1 and a dilation of A=5 (representing angular resolution) was used. A zero-padding was also applied to maintain the output's spatial size identical to the input MacPI. As illustrated in Figure 2, using light purple color, this filter processes only spatial information.

3.2.2 Epipolar Feature Extractor

To understand the twisted relationship between spatial and angular information, we extract features from horizontal and vertical epipolar lines. Therefore, we use special linear kernels to process horizontal and vertical slices on the MacPI. Specifically, we utilize convolution filters with a kernel size of 1×(A+2), (A+2)×1 and a zero-padding =(A+2)//2 to ensure the output has the same spatial size as the input MacPI, as illustrated in Figure 2, using red and green colors for vertical and horizontal extractors, respectively. In addition, the value of epipolar kernel length =A+2 was chosen to increase the receptive field of the network, and it was empirically proven that this value achieves the best performance, as demonstrated in the ablation study.

Different from the spatial feature extractor that processes two spatial dimensions, each epipolar feature extractor processes one spatial dimension and another angular dimension. To increase the interaction between different dimensions processes by the horizontal and vertical extractors, we used the horizontal and vertical epipolar feature extractors interchangeably to achieve maximum interaction and for a better understanding of the spatial-angular relationship.

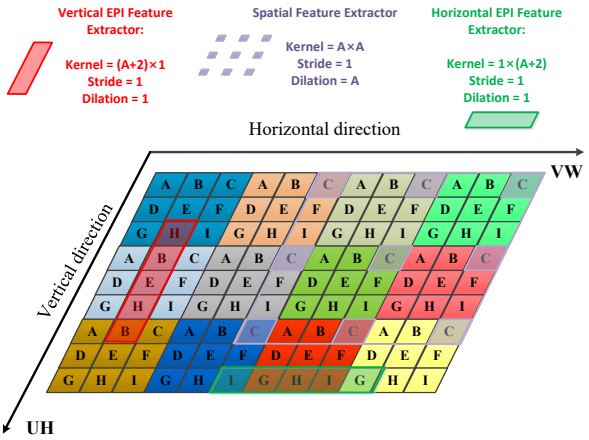

Figure 2. A visual example of the epipolar and spatial feature extractors. We use LF of size U=V=3, H=3, and W=4, where distinct macro-pixels are painted with different backdrop colors, and different characters represent pixels from different views. In particular, spatial feature extractors process pixels from the same views, while horizontal and vertical EPI feature extractors process horizontal and vertical epipolar lines, respectively.

3.3. Network Design

3.3.1 Overview

The architecture of the proposed network is shown in Figure 3a. We feed an LR input $L_{LR} \in R^{UH\times VW}$ to the network to generate an HR output $L_{HR} \in R^{\alpha UH\times \alpha VW}$. Following the SOTA methods [1, 23, 24], 1) We train the proposed network with upsampling factor α=2, 4. 2) We use the Y channel only to train the network after converting input LFs from RGB into YCbCr, while the Cb and Cr components are upsampled bicubically. 3) We used the same public datasets for training and testing (i.e., EPFL [19], HCInew [18], HCIold [20], INRIA [21], STFgantry [22]). 4) We employed PSNR and SSIM as quantitative performance measures.

We adopted the residual-in-residual structure [1, 32, 33], with N residual groups (Extract-Group), each containing N residual blocks (Extract-block), as shown in Figure 3a. First, as mentioned earlier, we convert the input LF organized as an array of SAIs into a MacPI to be processed using different feature extractors. Then, we extract initial features to be fed to the network using a single convolution layer. N-cascaded residual groups process these initial features to generate deep features and then convert them to an array of SAIs similar to the input. Next, we upsample the learnt deep features represented by an array of SAIs to increase their spatial resolution from (H, W) into (αH, αW). Fi-

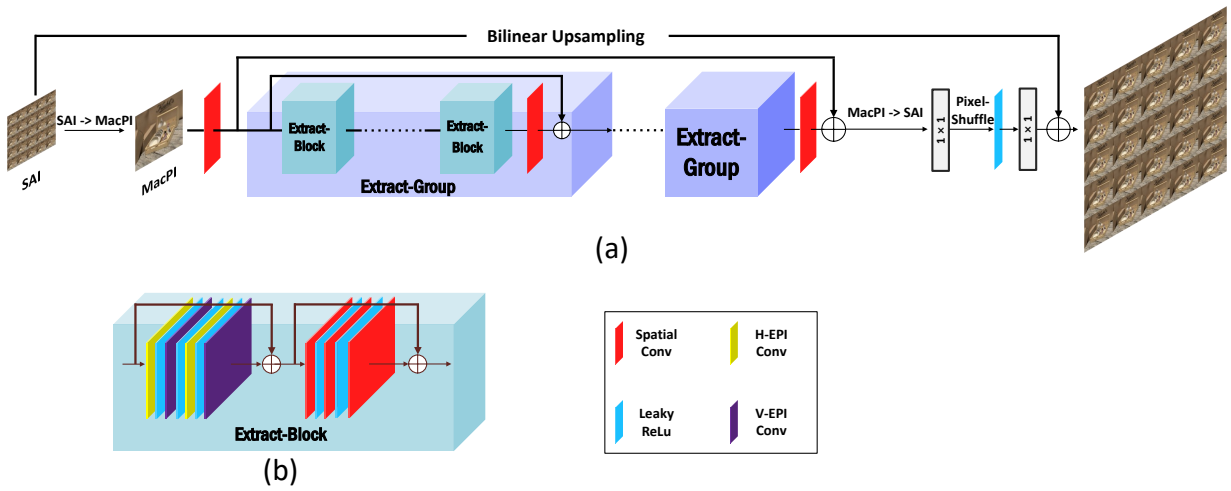

Figure 3. An overview of our proposed network.

nally, the upsampled image is added to the bilinear upsampled input to generate an output LF image.

3.3.2 Extract-Block

As illustrated in Figure 3b, the basic module in which spatial and epipolar information is processed sequentially is the Extract-Block. First, a residual epipolar block with four extractors incorporates the epipolar information. In our implementation, we use interchangeably two horizontal and two vertical epipolar feature extractors to understand the spatial-angular relationship better. Then, the epipolar extracted features are fed to the residual spatial block with three extractors incorporating the spatial information.

3.3.3 Spatial Upsampling

To increase the spatial resolution of the learnt deep features from (H, W) into (αH, αW), we follow [1, 15] and use the same layer consisting of two 1×1 convolutions with a pixel shuffler and a Leaky ReLu in between. The first convolution layer increases the features' depth from C to $\alpha^2 C$, and then the pixel shuffler reorders extended features from (H, W, $\alpha^2 C$) to (αH, αW, C). Finally, the last convolution squeezes the depth from C to 1. This upsampled image is then added to the bilinear upsampled input to generate an output LF image.

3.4. Training Details

We used five public datasets for training and testing with angular resolution 9×9 [18–22]. We angularly cropped the middle 5×5 SAIs, downsampled them bicubically, trimmed them into batches with 32×32 size and augmented them using multi-rotation and flipping. Finally, we trained the proposed network using L1 loss and optimized it utilizing the Adam method with the default parameters [34]. We implemented our network using Pytorch on a PC with Nvidia RTX 3090 GPU for 50 epochs, starting with a learning rate of 2×10^{-4} and halved every 15 epochs.

4. Experiments

4.1. Comparisons With SOTA Methods

We compare our method to 14 SOTA methods, including three single image super-resolution methods (i.e., VDSR [35], EDSR [36], RCAN [32]) and eleven LFSR methods (i.e., resLF [10], LFSSR [13], LF-ATO [12], LF-InterNet [15], LF-DFnet [14], MEG-Net [11], LF-IINet [16], DPT [17], DistgSSR [1], LFT [23], EPIT [24]).

1) Quantitative Results: We present quantitative results achieved by our method and other SOTA methods in Tables 1,2. We compare two model variants with the SOTA methods to validate our proposed method. The primary model (Ours) consists of eight residual groups containing eight blocks. We used this model to participate in the NITRE 2023 challenge [25]. In contrast, the second model (Ours-S) is a light version of the primary model consisting of five residual groups, each containing five residual blocks. Both models (Ours and Ours-S) have the same number of channels $=64$. Except for the STFgantry dataset on $2\times$ LFSR (which has more significant disparity variations), both models achieve competitive PSNR and SSIM results.

2) Qualitative Results: Figure 4 shows qualitative results on real-world LF scenes [38]. In addition, Figure 5 shows more qualitative results with SOTA methods for $2\times$ and $4\times$ LFSR on synthetic LF scenes. Our proposed method can preserve the textures and details in the super-resolved im-

Methods	Param.	2× EPFL	HCInew	HCIold	INRIA	STFgantry
Bicubic	-	29.74	31.89/.936	37.69/.979	31.33/.958	31.06/.950
VDSR [35]	0.66M	32.50/.960	34.37/.956	40.61/.987	34.43/,974	35.54/.979
EDSR [36]	38.6M	33.09/.963	34.83/.959	41.01/.987	34.97/.976	36.29/.982
RCAN [32]	15.3M	33.16/.963	34.98/.960	41.05/.988	35.01/.977	36.33/.983
resLF [10]	7.98M	33.62/.971	36.69/.974	43.42/.993	35.39/.980	38.36/.990
LFSSR [13]	0.88M	33.68/.974	36.81/.975	43.81/.994	35.28/.983	37.95/.990
LF-ATO [12]	1.22M	34.27/.976	37.24/.977	44.20/.994	36.15/.984	39.64/.993
LFInterNet [15]	5.04M	34.14/.976	37.28/.976	44.45/**.995**	35.80/.984	38.72/.991
LF-DFnet [14]	3.94M	34.44/.976	37.44/.977	44.23/.994	36.36/.984	39.61/.993
MEG-Net [11]	1.69M	34.30/.977	37.42/.978	44.08/.994	36.09/.985	38.77/.992
LF-IINet [16]	4.84M	34.68/.977	37.74/.979	44.84/**.995**	36.57/.985	39.86/.994
DPT [17]	3.73M	34.48/.976	37.35/.977	44.31/.994	36.40/.984	39.52/.993
DistgSSR [1]	3.53M	34.81/.979	37.96/.980	44.94/**.995**	36.59/.986	40.40/.994
LFT [23]	1.11M	34.80/.978	37.84/.979	44.52/**.995**	36.59/.986	40.51/.994
EPIT [24]	1.42M	34.83/.978	38.23/.981	45.08/**.995**	36.67/.985	**42.17/.996**
Ours-S	5.87M	35.16/.980	38.27/.980	45.04/**.995**	36.93/.987	40.77/.995
Ours	14.77M	**35.88/.984**	**38.57/.982**	**45.18/.995**	**37.33/.988**	41.41/.995

Table 1. Quantitative comparison for 2× LFSR. The best results are **bolded**.

Methods	Param.	4× EPFL	HCInew	HCIold	INRIA	STFgantry
Bicubic	-	25.14/.832	27.61/.852	32.42/.934	26.82/.887	25.93/.845
VDSR [35]	0.66M	27.25/.878	29.31/.882	34.81/.952	29.19/.920	28.51/.901
EDSR [36]	38.9M	27.84/.885	29.60/.887	35.18/.954	29.66/.926	28.70/.907
RCAN [32]	15.4M	27.88/.886	29.63/.889	35.20/.955	29.76/.928	28.90/.913
resLF [10]	8.64M	28.27/.904	30.73/.911	36.71/.968	30.34/.941	30.19/.937
LFSSR [13]	1.77M	28.27/.912	30.72/.915	36.70/.970	30.31/.947	30.15/.943
LF-ATO [12]	1.36M	28.52/.912	30.88/.914	37.00/.970	30.71/.948	30.61/.943
LFInterNet [15]	5.48M	28.67/.916	30.98/.916	37.11/.972	30.64/.949	30.53/.941
LF-DFnet [14]	3.99M	28.77/.917	31.23/.920	37.32/.972	30.83/.950	31.15/.949
MEG-Net [11]	1.77M	28.74/.916	31.10/.918	37.28/.972	30.66/.949	30.77/.945
LF-IINet [16]	4.88M	29.11/.919	31.36/.921	37.62/.973	31.08/.952	31.21/.950
DPT [17]	3.78M	28.93/.917	31.19/.919	37.39/.972	30.96/.950	31.14/.949
DistgSSR [1]	3.58M	28.99/.920	31.38/.922	37.56/.973	30.99/.952	31.65/.954
LFT [23]	1.16M	29.25/.921	31.46/.922	37.63/.974	31.20/.952	31.86/.955
EPIT [24]	1.47M	**29.34**/.920	31.51/.923	37.68/.974	31.27/.953	32.18/.957
Ours-S	5.92M	**29.34**/.925	31.70/.925	37.97/.975	**31.36**/.955	32.20/.958
Ours	14.82M	29.33/**.927**	**31.80/.927**	**38.04/.976**	31.35/**.956**	**32.36/.960**

Table 2. Quantitative comparison for 4× LFSR. The best results are **bolded**.

ages and achieve competitive visual performance, obvious through the zoom-in regions.

3) Angular Consistency: Using the 4× output of different SOTA methods, we calculate the depth [37] (The quality of estimated depth depends on the angular consistency of each LFSR method), as shown in Figure 6. The technique employed for depth estimation uses a spinning parallelogram operator (SPO). SPO overcomes occlusion, noise, and outliers in LF data and achieves robust depth estimation even in challenging scenes [37]. The SPO is used to extract consistent and reliable depth information, and a multi-stage processing strategy is used for improved accuracy.

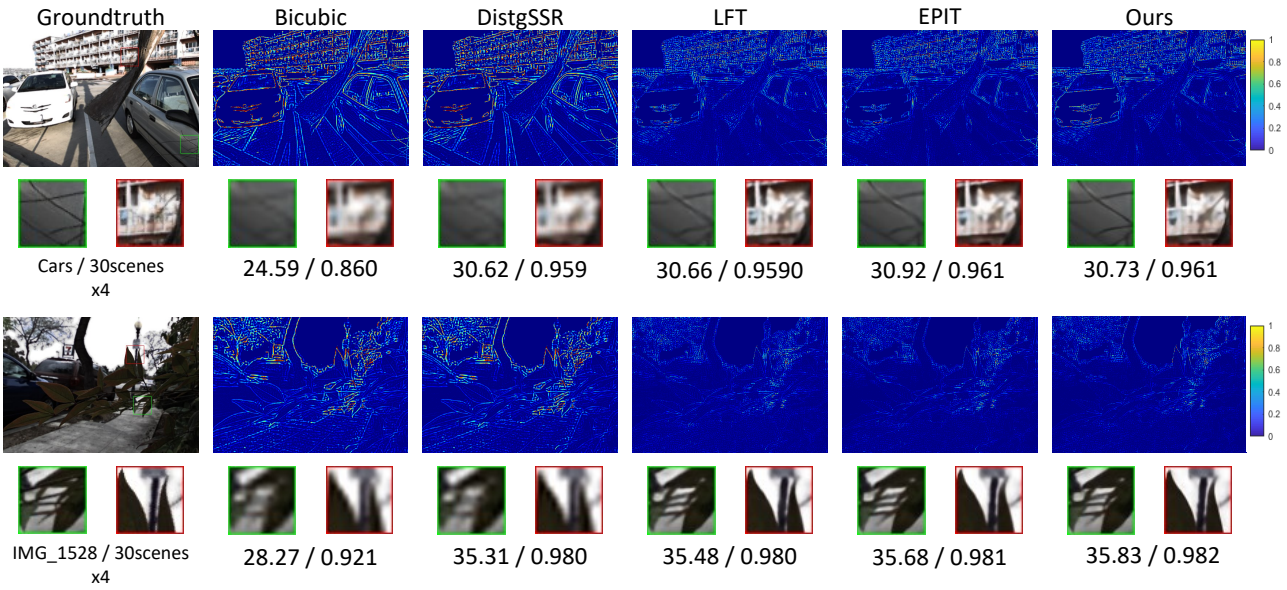

Figure 4. Visual results aquired by SOTA methods for $4\times$ LFSR on real-world LF scenes.

	Spatial	Epipolar	EPFL	HCInew	HCIold	INRIA	STFgantry
1	$\surd$		27.87/.8866	29.69/.8883	35.26/.9542	29.73/.9268	29.13/.9140
2		$\surd$	29.20/.9237	31.54/.9233	37.69/.9741	31.19/.9544	31.63/.9540
3	$\surd$	$\surd$*	28.73/.8451	30.54/.8197	36.23/.9146	30.70/.8867	30.37/.9047
4	$\surd$	$\surd$**	29.25/.8721	31.51/.8531	37.71/.9411	31.29/.9088	31.82/.9345
5	$\surd$	$\surd$	29.34/.9250	31.70/.9250	37.97/.9751	31.36/.9554	32.20/.9580

Table 3. Quantitative comparison between different variants of our proposed model to study the effect of each component. $\surd$* indicates only horizontal epipolar extractors. $\surd$** indicates separate horizontal and vertical epipolar extractors.

4.2. Ablation Studies

4.2.1 Effect of Each Component

We train different variants of our proposed model (the light version: Ours-S) to validate each component's effectiveness, as presented in Table 3 . We adjust the number of parameters of each model to be close to the main model (Ours-S). Model-1 uses only spatial feature extractors (similar to SISR models), as shown in Figure 7b. This model cannot incorporate angular information and provides the minimum PSNR and SSIM. In contrast, model-2 uses epipolar feature extractors only and achieves the second-best performance, as shown in Figure 7c. Results achieved by model-1 and model-2 highlight the effect of epipolar feature extractors to understand the spatial-angular relationship.

In model-3 and model-4, we compare different implementations of the residual epipolar block. Model-3 uses horizontal extractors only, as shown in Figure 7d, while model-4 uses separate horizontal and vertical residual epipolar blocks with a fusion layer to fuse the extracted features, as shown in Figure 7e. However, model-4 provides slightly better PSNR values than model-3 and is very close to the main model; both models cannot provide good SSIM values nor understand well the spatial-angular relationship.

4.2.2 Effect of Epipolar Kernel Length

We train different variants to validate the effect of the epipolar kernel length. Each model interchangeably contains two horizontal and two vertical epipolar feature extractors with the same kernel length, as presented in Table 4. The model performance increases by increasing the kernel length from A to A+2, as shown in model-1 and model-2, especially with datasets with larger disparity variations (i.e., STFgantry). Increasing the kernel length provides more information; hence, the network can better understand the spatial-angular relationship. However, with longer kernels, understating this relationship becomes more complex, and the performance starts to decrease, as shown in model-2 and model-3, when the kernel increases from A+2 to A+4.

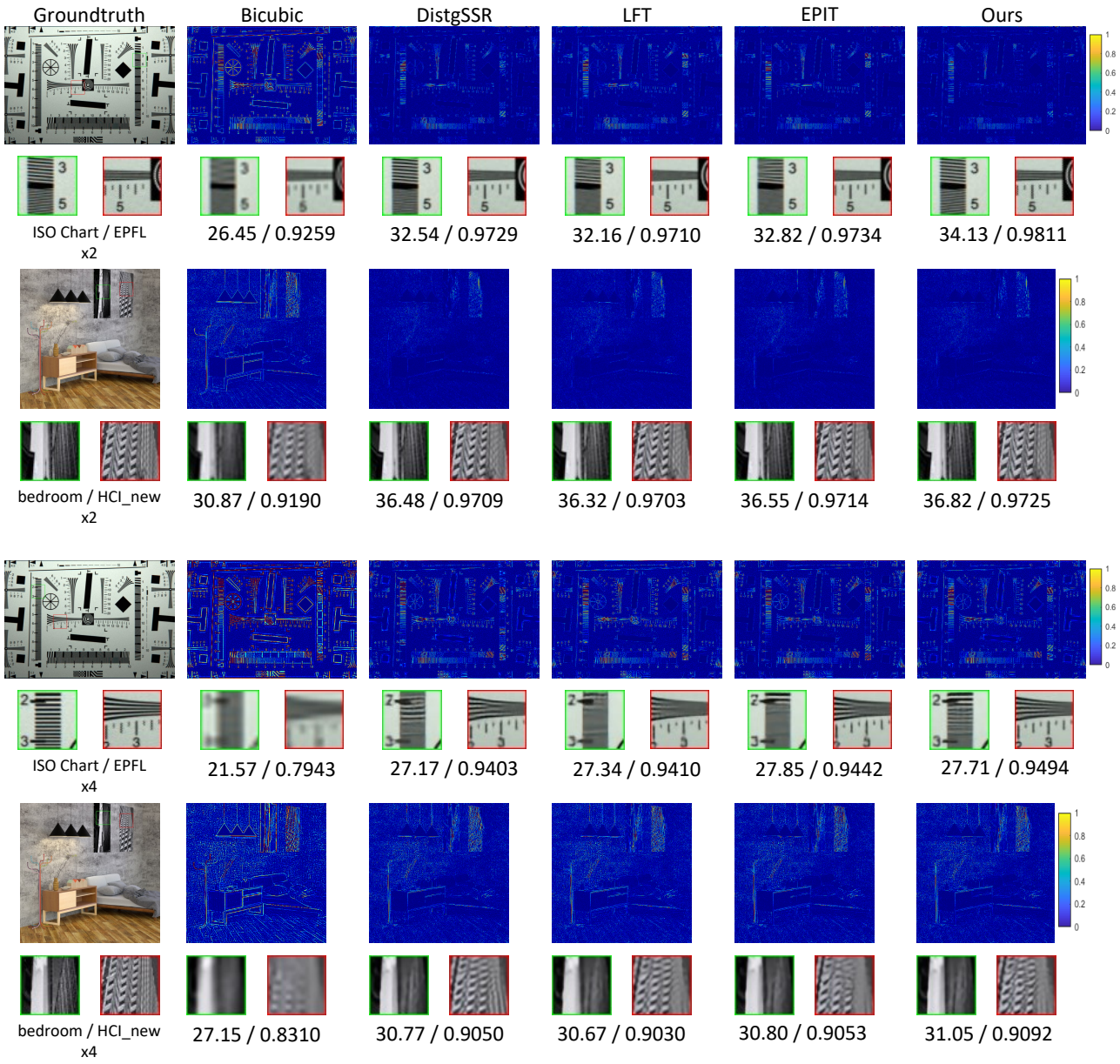

Figure 5. Visual results aquired by SOTA methods for $2\times$ and $4\times$ LFSR on synthetic LF scenes.

	Kernel length	EPFL	HCInew	HCIold	INRIA	STFgantry
1	$A = 5$	29.33/.9245	31.63/.9241	37.74/.9745	31.37/.9548	31.81/.9555
2	$A + 2 = 7$	29.34/.9250	31.70/.9250	37.97/.9751	31.36/.9554	32.20/.9580
3	$A + 4 = 9$	29.23/.8737	31.62/.8549	37.90/.9429	31.29/.9104	31.92/.9369

Table 4. Quantitative comparison between variants with different epipolar kernel length values.

4.3. NTIRE LFSR Challenge 2023

This challenge uses the same five public LF datasets for training [18–22], with new datasets for validation and testing, each containing 16 synthetic and 16 real-world images [25]. We compare our method to 10 other teams participating in the final test phase, producing higher PSNR than

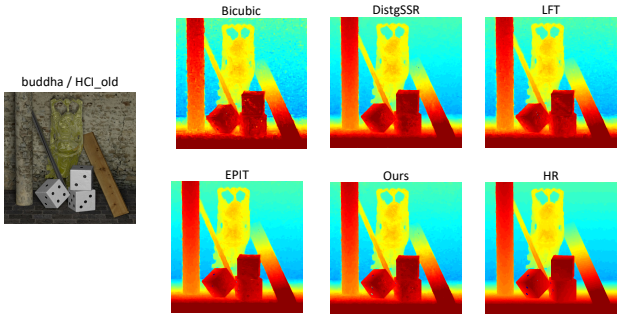

Figure 6. Depth estimated using SPO [37] on results aquired by SOTA methods for 4× LFSR.

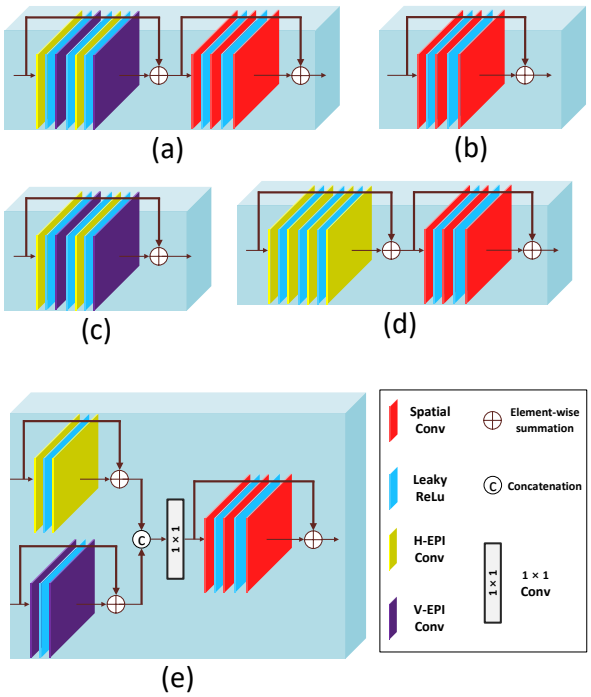

Figure 7. Different implementations of the Extract-Block. a) Original module using epipolar and spatial blocks. b) Using spatial block only. c) Using epipolar block only. d) Original block with horizontal epipolar only. e) Original block with separate horizontal and vertical epipolar blocks.

the baseline method (LF-InterNet [15]).

Tables 5 and 6 present quantitative comparisons between our results (ranked 10th on the test set and 6th on the validation set) and other teams on the validation and testing datasets, respectively.

5. Conclusion

This research presents a simple yet effective method for LF Image SR. Within a residual-in-residual framework, we suggested spatial and interchangeable epipolar feature extractors. The epipolar and spatial information are processed sequentially through residual epipolar and residual spatial blocks inside the Extract-Block (the basic module of our network). Our technique produced competitive results in the NTIRE Light Field Image Super-Resolution Challenge 2023: our model was ranked 10th on the test set and 6th on the validation set. Furthermore, our model produced quantitative and qualitative competitive outcomes compared to the SOTA approaches.

	Team	Average	Lytro	Synthetic
1	OpenMeow	32.71/.9496	33.36/.9562	32.07/.9430
2	VIDAR	32.54/.9494	33.24/.9568	31.85/.9419
3	DMLab	32.43/.9485	33.24/.9559	31.62/.9410
4	BNU-AI-TRY	32.29/.9468	32.96/.9539	31.63/.9396
5	IIR-Lab	32.24/.9465	32.84/.9529	31.64/.9402
6	**Ours**	**32.13/.9464**	**32.70/.9533**	**31.55/.9395**
7	HawkeyeGroup	32.13/.9463	32.86/.9543	31.40/.9383
8	Insis	32.12/.9455	32.86/.9526	31.39/.9383
9	BIT912	32.05/.9449	32.76/.9528	31.35/.9371
10	SHU-IVIPLab	32.01/.9442	32.69/.9517	31.32/.9366
11	LFSRgdutteam	31.83/.9431	32.53/.9508	31.13/.9354

Table 5. Quantitative comparison on the validation set of the NTIRE LFSR challenge 2023 [25].

	Team	Average	Lytro	Synthetic
1	OpenMeow	30.66/.9314	30.82/.9475	30.51/.9152
2	DMLab	30.64/.9318	30.92/.9489	30.35/.9146
3	VIDAR	30.56/.9323	30.67/.9491	30.45/.9154
4	IIR-Lab	30.38/.9285	30.56/.9450	30.20/.9119
5	Insis	30.35/.9287	30.56/.9458	30.15/.9117
6	BNU-AI-TRY	30.13/.9290	29.97/.9453	30.29/.9126
7	BIT912	30.11/.9293	30.10/.9465	30.13/.9120
8	HawkeyeGroup	30.06/.9285	29.99/.9447	30.13/.9124
9	SHU-IVIPLab	29.90/.9265	29.78/.9433	30.01/.9096
10	**Ours**	**29.85/.9279**	**29.64/.9447**	**30.06/.9111**
11	LFSRgdutteam	29.83/.9262	29.64/.9422	30.01/.9103

Table 6. Quantitative comparison on the testing set of the NTIRE LFSR challenge 2023. [25].

Acknowledgement

This work was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education under Grant 2023R1A2C1006944 and 2020R1I1A3A04037680.

References

[1] Yingqian Wang, Longguang Wang, Gaochang Wu, Jungang Yang, Wei An, Jingyi Yu, and Yulan Guo. Disentangling light fields for super-resolution and disparity estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(1):425–443, 2022. 1, 2, 3, 4, 5

[2] Yingqian Wang, Longguang Wang, Zhengyu Liang, Jungang Yang, Wei An, and Yulan Guo. Occlusion-aware cost constructor for light field depth estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 19809–19818, 2022. 1

[3] Keren Fu, Yao Jiang, Ge-Peng Ji, Tao Zhou, Qijun Zhao, and Deng-Ping Fan. Light field salient object detection: A review and benchmark. *Computational Visual Media*, 8(4):509–534, 2022. 1

[4] Yi Zhang, Geng Chen, Qian Chen, Yujia Sun, Yong Xia, Olivier Deforges, Wassim Hamidouche, and Lu Zhang. Learning synergistic attention for light field salient object detection. *arXiv preprint arXiv:2104.13916*, 2021. 1

[5] Jiwan Hur, Jae Young Lee, Jaehyun Choi, and Junmo Kim. I see-through you: A framework for removing foreground occlusion in both sparse and dense light field images. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 229–238, 2023. 1

[6] Yingqian Wang, Tianhao Wu, Jungang Yang, Longguang Wang, Wei An, and Yulan Guo. Deoccnet: Learning to see through foreground occlusions in light fields. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 118–127, 2020. 1

[7] Hao Zhu, Mantang Guo, Hongdong Li, Qing Wang, and Antonio Robles-Kelly. Revisiting spatio-angular trade-off in light field cameras and extended applications in super-resolution. *IEEE transactions on visualization and computer graphics*, 27(6):3019–3033, 2019. 1

[8] Gaochang Wu, Belen Masia, Adrian Jarabo, Yuchen Zhang, Liangyong Wang, Qionghai Dai, Tianyou Chai, and Yebin Liu. Light field image processing: An overview. *IEEE Journal of Selected Topics in Signal Processing*, 11(7):926–954, 2017. 1

[9] Yangling Chen, Shuo Zhang, Song Chang, and Youfang Lin. Light field reconstruction using efficient pseudo 4d epipolar-aware structure. *IEEE Transactions on Computational Imaging*, 8:397–410, 2022. 1

[10] Shuo Zhang, Youfang Lin, and Hao Sheng. Residual networks for light field image super-resolution. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 11046–11055, 2019. 1, 2, 4, 5

[11] Shuo Zhang, Song Chang, and Youfang Lin. End-to-end light field spatial super-resolution network using multiple epipolar geometry. *IEEE Transactions on Image Processing*, 30:5956–5968, 2021. 1, 2, 4, 5

[12] Jing Jin, Junhui Hou, Jie Chen, and Sam Kwong. Light field spatial super-resolution via deep combinatorial geometry embedding and structural consistency regularization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2260–2269, 2020. 1, 2, 4, 5

[13] Henry Wing Fung Yeung, Junhui Hou, Xiaoming Chen, Jie Chen, Zhibo Chen, and Yuk Ying Chung. Light field spatial super-resolution using deep efficient spatial-angular separable convolution. *IEEE Transactions on Image Processing*, 28(5):2319–2330, 2018. 1, 2, 4, 5

[14] Yingqian Wang, Jungang Yang, Longguang Wang, Xinyi Ying, Tianhao Wu, Wei An, and Yulan Guo. Light field image super-resolution using deformable convolution. *IEEE Transactions on Image Processing*, 30:1057–1071, 2020. 1, 2, 4, 5

[15] Yingqian Wang, Longguang Wang, Jungang Yang, Wei An, Jingyi Yu, and Yulan Guo. Spatial-angular interaction for light field image super-resolution. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXIII 16*, pages 290–308. Springer, 2020. 1, 2, 4, 5, 8

[16] Gaosheng Liu, Huanjing Yue, Jiamin Wu, and Jingyu Yang. Intra-inter view interaction network for light field image super-resolution. *IEEE Transactions on Multimedia*, 2021. 1, 2, 4, 5

[17] 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. 1, 4, 5

[18] Katrin Honauer, Ole Johannsen, Daniel Kondermann, and Bastian Goldluecke. A dataset and evaluation methodology for depth estimation on 4d light fields. In *Computer Vision–ACCV 2016: 13th Asian Conference on Computer Vision, Taipei, Taiwan, November 20-24, 2016, Revised Selected Papers, Part III 13*, pages 19–34. Springer, 2017. 1, 3, 4, 7

[19] Martin Rerabek and Touradj Ebrahimi. New light field image dataset. In *8th International Conference on Quality of Multimedia Experience (QoMEX)*, number CONF, 2016. 1, 3, 4, 7

[20] Sven Wanner, Stephan Meister, and Bastian Goldluecke. Datasets and benchmarks for densely sampled 4d light fields. In *VMV*, volume 13, pages 225–226, 2013. 1, 3, 4, 7

[21] Mikael Le Pendu, Xiaoran Jiang, and Christine Guillemot. Light field inpainting propagation via low rank matrix completion. *IEEE Transactions on Image Processing*, 27(4):1981–1993, 2018. 1, 3, 4, 7

[22] Vaibhav Vaish and Andrew Adams. The (new) stanford light field archive. *Computer Graphics Laboratory, Stanford University*, 6(7):3, 2008. 1, 3, 4, 7

[23] Zhengyu Liang, Yingqian Wang, Longguang Wang, Jungang Yang, and Shilin Zhou. Light field image super-resolution with transformers. *IEEE Signal Processing Letters*, 29:563–567, 2022. 1, 2, 3, 4, 5

[24] 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. *arXiv preprint arXiv:2302.08058*, 2023. 1, 2, 3, 4, 5

[25] Yingqian Wang, Longguang Wang, Zhengyu Liang, Jungang Yang, Radu Timofte, and Yulan Guo. Ntire 2023 challenge on light field image super-resolution. In *CVPRW*, 2023. 2, 4, 7, 8

[26] 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, 2013. 2

[27] 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

[28] Reuben A Farrugia, Christian Galea, and Christine Guillemot. Super resolution of light field images using linear subspace projection of patch-volumes. *IEEE Journal of Selected Topics in Signal Processing*, 11(7):1058–1071, 2017. 2

[29] Martin Alain and Aljosa Smolic. Light field super-resolution via lfbm5d sparse coding. In *2018 25th IEEE international conference on image processing (ICIP)*, pages 2501–2505. IEEE, 2018. 2

[30] 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

[31] Marc Levoy and Pat Hanrahan. Light field rendering. In *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*, pages 31–42, 1996. 2

[32] Yulun Zhang, Kunpeng Li, Kai Li, Lichen Wang, Bineng Zhong, and Yun Fu. Image super-resolution using very deep residual channel attention networks. In *Proceedings of the European conference on computer vision (ECCV)*, pages 286–301, 2018. 3, 4, 5

[33] Ahmed Salem, Hatem Ibrahem, and Hyun-Soo Kang. Light field reconstruction using residual networks on raw images. *Sensors*, 22(5):1956, 2022. 3

[34] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 4

[35] Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee. Accurate image super-resolution using very deep convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1646–1654, 2016. 4, 5

[36] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. Enhanced deep residual networks for single image super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 136–144, 2017. 4, 5

[37] Shuo Zhang, Hao Sheng, Chao Li, Jun Zhang, and Zhang Xiong. Robust depth estimation for light field via spinning parallelogram operator. *Computer Vision and Image Understanding*, 145:148–159, 2016. 5, 8

[38] Nima Khademi Kalantari, Ting-Chun Wang, and Ravi Ramamoorthi. Learning-based view synthesis for light field cameras. *ACM Transactions on Graphics (TOG)*, 35(6):1–10, 2016. 4