electronics

Article

MFSR: Light Field Images Spatial Super Resolution Model Integrated with Multiple Features

Jianfei Zhou and Hongbing Wang *

School of Computer and Communication Engineering, University of Science and Technology Beijing, Beijing 100083, China
* Correspondence: wanghongbing0816@163.com

Abstract: Light Field (LF) cameras can capture angular and spatial information simultaneously, making them suitable for a wide range of applications such as refocusing, disparity estimation, and virtual reality. However, the limited spatial resolution of the LF images hinders their applicability. In order to address this issue, we propose an end-to-end learning-based light field super-resolution (LFSR) model called MFSR, which integrates multiple features, including spatial, angular, epipolar plane images (EPI), and global features. These features are extracted separately from the LF image and then fused together to obtain a comprehensive feature using the Feature Extract Block (FE Block) iteratively. Gradient loss is added into the loss function to ensure that the MFSR has good performance for LF images with rich texture. Experimental results on synthetic and real-world datasets demonstrate that the proposed method outperforms other state-of-the-art methods, with a peak signal-to-noise ratio (PSNR) improvement of 0.208 dB and 0.274 dB on average for the 2× and 4× super-resolution tasks, and structural similarity (SSIM) of both improvements of 0.01 on average.

Keywords: macro-pixel images; super-resolution; global feature; gradient loss

check for updates

Citation: Zhou, J.; Wang, H. MFSR: Light Field Images Spatial Super Resolution Model Integrated with Multiple Features. *Electronics* **2023**, *12*, 1480. https://doi.org/10.3390/electronics12061480

Academic Editors: Beiji Zou, Xiaoyan Kui and Weixin Si

Received: 10 February 2023
Revised: 15 March 2023
Accepted: 17 March 2023
Published: 21 March 2023

1. Introduction

LF cameras have the unique ability to capture both the intensity and direction of light rays, making it possible to record 3D geometry in a convenient and efficient manner. By encoding 3D scene information into 4D LF images (2D for spatial dimensions and 2D for angular dimensions), LF cameras enable a wide range of applications such as post-capture refocusing [1,2], depth sensing [3,4], saliency detection [5,6], and de-occlusion [7,8]. However, there is a trade-off between spatial and angular resolution, with LF cameras providing either dense angular samplings with low image resolution or high-resolution (HR) sub-aperture images (SAI) with sparse angular samplings. To overcome this limitation, researchers focus on reconstructing high-resolution LF images from low-resolution(LR) ones, a process known as LF image spatial super-resolution (SR). Extracting and utilizing the different features in the LF images is a key topic in LF image spatial SR. In this paper, we focus on extracting multiple features of LF to achieve LF image spatial SR.

Conventional 2D images provide a single-view description, while 4D LF captures multiple views of a scene. In order to achieve light field super-resolution, researchers have typically applied 2D image SR models to each sub-aperture image, but this approach fails to utilize the angular features of LF. To address this issue, many researchers have focused on exploiting multiple features of different SAI to improve super-resolution reconstruction quality. Some methods, such as [9,10], treat the entire 4D LF as a single entity to exploit angle and spatial features but ignore LF sub-space features. Others, like [11], disentangle LF sub-space features for LF super-resolution, they involve global features, which refer to the overall embedding of an LF image, such as LF texture, geometry embedding, and structural embedding. We propose an end-to-end light field image super-resolution model MFSR that integrates multiple features, including spatial, angular, EPI, and global features.

The challenge of super-resolving rich texture areas is not limited to LFSR but also affects Single Image Super-Resolution (SISR). To address this issue, Jiang et al. proposed a focal frequency loss function [12] to enhance image reconstruction. Wang et al. recently introduced DPT [13], which preserves detailed texture information in LF images by including a gradient image as an independent branch in the network. The current LFSR methods have limited reconstruction performance in areas with rich textures. To address this bottleneck, a gradient loss is added to guide the model in learning these rich texture features in LF images, which ensures our MFSR has good performance for LF images with rich texture.

Our proposed method, MFSR, brings several innovations to the light field spatial super-resolution. First, we introduce a flexible and efficient end-to-end LFSR model that achieves state-of-the-art (SOTA) results across different LF image datasets. Second, we build upon multiple sub-space feature extraction by proposing a global feature extract block and fusing multiple features to enhance LFSR performance. Third, the gradient loss on Sobel [14] was applied to ensure MFSR has outperformance for LF image with rich texture. Our experiments, which include both synthetic and real-world scenarios, confirm the effectiveness of our MFSR model. The synthetic scenarios refer to the creation of synthetic LF data by computer science, such as the HCI Synthetic Light Field Dataset [15] and EPFL Light Field Dataset [16]. We demonstrate through experiments that our model MFSR can handle super-resolution tasks at different scales and achieve better results than the baseline. Our results indicate that our model outperforms the state-of-the-art LFSR model, where the PSNR results are improved by 0.208 dB and 0.274 dB on average in ×2 and ×4 light super-resolution tasks.

2. Related Works

The primitive classic non-learning-based methods utilize projection and optimization techniques to super-resolve the SAI, relying on geometric [17] and mathematical [18] modeling of the 4D light field structure. These classic LFSR methods establish a mathematical model through relation among different views and solve the mathematical model by information including disparities and geometric. Wanner et al. [15,19] took the lead to estimate disparity information of light field by an algorithm based on EPIs and then obtain EPIs after super-resolution by corresponding interpolation through disparities, and then transformed LF EPIs into SAIs to obtain super-resolution results. Mitra et al. [20] proposed an LFSR algorithm based on disparity estimation through pixel block in the light field based on Gaussian Mixed Model (GMM). There are also algorithms [21,22] able to compensate and recover view by matching relative pixels from other views. Rossi et al. [23] solved the problem by regarding LFSR as a global optimization question through an image optimization model. Farrugia et al. [24] broke HR and LR image couples into several sub-spaces and obtained light field spatial images through Principal Component Analysis (PCA). All the above methods need view disparity to calculate the transmission matrix between views in the light field. There is still a shortcoming in the aspect of digging out relationships among views and spatial information.

With the application of deep learning in computer vision, convolution neural network (CNN) has represented excellent performance in the task of computer vision. A series of algorithms based on deep learning for image super-resolution are proposed including SRCNN [25], SRGAN [26], VDSR [27], EDSR [28], RCAN [29]. Compared to classic non-learning-based methods, these methods obtain higher PSNR and SSIM scores in solving SISR. Since the learning-based SISR methods have made much progress, several researchers begin to research light field super-resolution (LFSR) based on deep convolution neural networks.

Recently, light field spatial super-resolution methods based on deep learning have been proposed. These methods learn angular and spatial features of LF images in the training process and represent wonderful performance in the aspect of qualitative and quantitative assessment. Yoon et al. [30] initially utilized a convolution neural network for spatial and angular super-resolution of LF images. In their work, every SAI in the LF is up-sampling by a single image super-resolution algorithm and its details are enhanced with a combination of 4D LF quality. Furthermore, images of new visual angles between adjacent views are generated by an angular super-resolution network. Zhang et al. proposed a light field images super-resolution method with the utilization of residual convolution neural network, ResLF [10], learning residual information in horizontal, vertical, and diagonal directions in the SAI array and applying to supplement the information in a high frequency of target view to super-resolution reconstruction. Jin et al. proposed an all-to-one light field image super-resolution architecture, LF-ATO [31], and propelled structure consistency regularization to protect disparity structure between reconstitution views. Coming up with LF-InterNet [32], Wang et al. extracted angular and spatial features in the light field images for assistance in the reconstruction of HR light field images through macro-pixel images. The experiment indicates that this method represents excellent performance in both vision and evaluation. Wang et al. subsequently applied deformable convolution to light field images spatial super-resolution algorithm and put forward LF-DFNet [33] which represents further promotion in PSNR score. Zhang et al. proposed MEG-Net [9] which stacks SAIs in different directions and utilizes 3D convolution to extract EPI features in different directions. Wang et al. proposed DistgSSR [11] to propel LFSR through decoupling different subspace features in the LF images which integrate with spatial, angular, and EPI features. Recently, Wang et al. came up with DPT [13] which initially applies the transformer structure to LFSR and it retains detailed texture information in the light field images through integration with detailed gradient information.

Previous LFSR algorithms typically treated spatial, angular, EPI, and global features as separate entities and did not consider integrating global features into the super-resolution process. In contrast, we propose an LFSR method to extract and integrate these features to improve LFSR performance. To preserve the detailed texture information in light field images during super-resolution, we apply the gradient loss to aid in the learning process and further enhance the performance of rich texture with LF images.

3. Methods

In recent work, Wang et al. proposed DistgSSR [11] to explore disentangled features for LF super-resolution tasks. While achieving state-of-the-art performance, we believe that there is still room for improvement in the architecture of DistgSSR [11]. Specifically, DistgSSR [11] maps the 4D LF to a 2D subspace for feature learning and fusion, but it cannot extract global features. In our method, we introduce a global feature extraction block based on spatial, angular, and EPI features to learn the inner features of the LF. Furthermore, our approach employs the gradient loss on Sobel [14] to ensure the performance of LF images with rich textures. We will elaborate on our super-resolution algorithm in detail in the following sections.

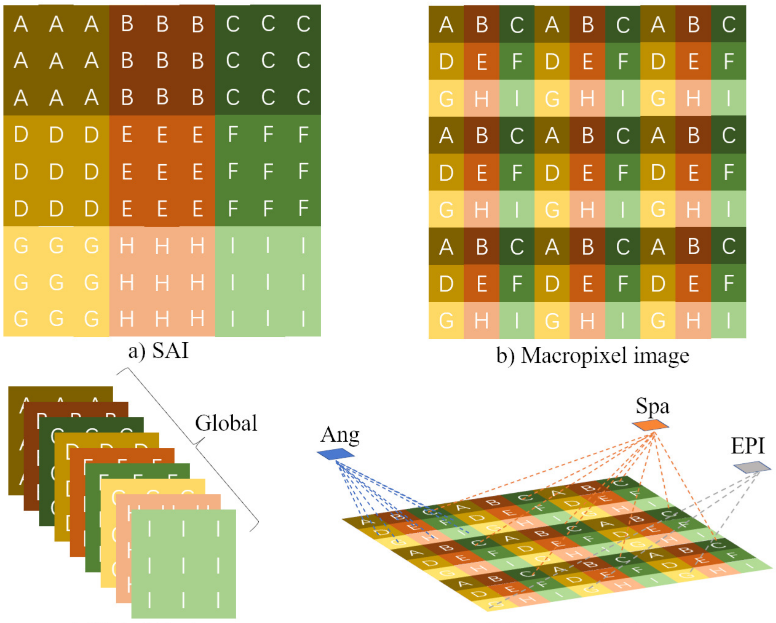
3.1. Light Field Representation

In the context of a light field image, 4D light field data can be represented as $L(u, v, h, w) \in R^{U \times V \times H \times W}$, where u and v correspond to the angular dimensions and h and w correspond to the spatial dimensions. However, due to the inconvenience of processing 4D data, current LFSR research often organizes light field images into SAI arrays or macro-pixel images. The four main types of LF features are spatial, angular, EPI, and global features. Spatial features are inherent to individual views and can be extracted similarly to traditional 2D images. Angular features capture the relationships between different views and can be used to further improve super-resolution performance. EPI features describe the relationship between space and angle in a certain direction, serving as a bridge between spatial and angular features. Global features refer to the overall embedding of an LF image, such as its texture, geometry embedding, and structural embedding. Our global features extractor uses 3D convolution to extract overall embedding from stacking SAI.

The SAI array provides a convenient visualization of the LF images where each sub-aperture image can be treated as a traditional 2D image. For example, Figure 1 illustrates a toy example of the SAI representation method. In the SAI array, LF images are arranged as a $U \times V$ image matrix with a spatial resolution of $H \times W$. Each SAI at a specific coordinate (u, v) is represented as $L(u, v, :, :) \in R^{H \times W}$, which can be treated as a traditional 2D image and processed using 2D image algorithms. The initial LFSR models based on deep learning treat each SAI as a traditional 2D image and employ single-image super-resolution algorithms to upscale them, followed by a fine-tuning step to integrate angular information. This approach is effective in extracting spatial and global features but may fail to extract features that are hidden among different views.

In the SAI array of LF images, stacking all pixels with a direction, an EPI could be obtained. We stock images with the v-axis, then stacked EPI image space is $L(u, :, h, :) \in R^{V \times W}$, which could represent linear relation between different views in the v-axis. The EPI image stacked all pixels in a certain direction which could represent the relation between spatial and angle in this direction. EPI features are applied to LFSR methods [9,11] which promotes the performance of LFSR based on spatial and angular features. In reference [11], fusion with EPI feature could be applied to not only LFSR but also LF images disparity estimation and LF images angular super-resolution, which obtains improved performance compared to only considering spatial and angular features. This research demonstrates the importance of EPI features in LF image processing.

4D LF could be also organized into macro-pixel images $L(:, :, w, h) \in R^{U \times V}$, are a series of pixels with the spatial coordinates (w, h) recorded by different cameras at different locations, a toy example as shown in Figure 1b. The process of converting SAIs into macro-pixel images is to organize the pixels at the same spatial location in each SAI according to their different angular coordinates, and then organize the macro-pixels at different spatial positions into a complete LF micro-pixel image. For the macro-pixel LF images, spatial and angular feature extraction is convenient. LFSR methods [11,32,33] take the macro-pixel images as the input and separately extract angular and spatial features to obtain the SOTA performance of the LFSR task. The organization method of macro-pixel could be not only convenient for extraction of spatial and angular features but also get EPI features by the pixel stack with a certain direction.



A	A	A	B	B	B	C	C	C
A	A	A	B	B	B	C	C	C
A	A	A	B	B	B	C	C	C
D	D	D	E	E	E	F	F	F
D	D	D	E	E	E	F	F	F
D	D	D	E	E	E	F	F	F
G	G	G	H	H	H	I	I	I
G	G	G	H	H	H	I	I	I
G	G	G	H	H	H	I	I	I

a) SAI

A	B	C	A	B	C	A	B	C
D	E	F	D	E	F	D	E	F
G	H	I	G	H	I	G	H	I
A	B	C	A	B	C	A	B	C
D	E	F	D	E	F	D	E	F
G	H	I	G	H	I	G	H	I
A	B	C	A	B	C	A	B	C
D	E	F	D	E	F	D	E	F
G	H	I	G	H	I	G	H	I

b) Macropixel image

c) Global Feature

d) Subspace Feature

Figure 1. Different representations and different features of light field images, they should be listed as: (**a**) Light Field Sub-aperture image. (**b**) Macro-pixel Light Field Image. (**c**) Global Feature (**d**) Spatial, Angular and EPI Feature.

The aforementioned LF organization methods have their respective advantages. LF SAI can be processed similarly to traditional 2D images but is limited in its ability to extract angular features. EPIs can represent the 2D section in a direction of a 4D LF clip but only demonstrate the relation between different views in a direction, which means they can only represent 1D spatial and angle features of LF and cannot extract 2D angular, spatial, or global features. LF macro-pixel images cannot be processed by traditional image algorithms and are used as a means of mosaic, but they are convenient for angular features extracted by 2D convolution. Additionally, due to LF angular resolution being much less than spatial resolution, spatial features can be obtained using a 2D dilated convolution, making it convenient for spatial and angular feature extraction. Furthermore, LF EPI features can be conveniently extracted from micro-pixel images and SAI. Therefore, we utilize macro-pixel LF images as input for the MFSR network.

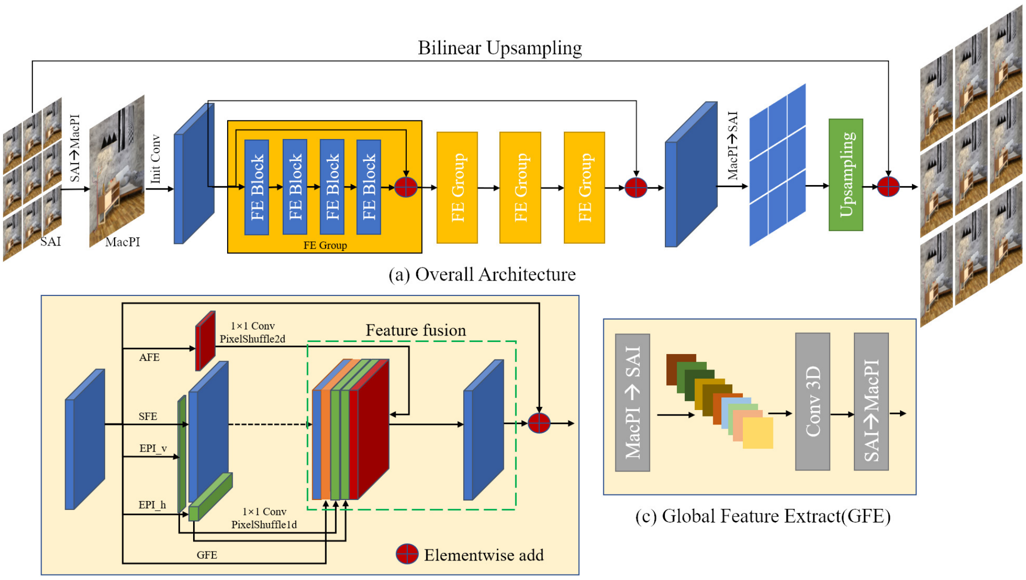
3.2. Network Design

The objective of the proposed MFSR network takes a low-resolution macro-pixel image with a size of $R^{U \times V \times H \times W}$ as input to produce a high-resolution sub-aperture image array with a size of $R^{U \times V \times \alpha H \times \alpha W}$. α represents the rate of spatial super resolution, U, V represents the angular resolution, and H, W represents spatial resolution. The MFSR network enters a low-resolution 4D light field image $I_{LR} \in R^{U \times V \times H \times W}$, and its corresponding feature extractions are separately extracted through spatial, angular, EPI, and global feature extractors. Then, the needed feature of super-resolution is obtained by integration with different extracted features. In the past light field feature extractor, the sub-aperture array is always utilized. Therefore, in the paper, the angular resolution of the 4D light field is set as $U = V = A$, and A means the angular resolution of the 4D light field.

3.2.1. Network Structure

The proposed network MFSR, as illustrated in Figure 2, takes light field sub-aperture images and converts them into macro-pixel images for initial convolution. The resulting features are then fed into four feature extraction blocks for spatial, angular, EPI, and global feature extraction. The core idea of the LFSR model based on deep learning is to learn multiple features from different angles and single views. In particular, EPI and angular features represent different angular information, spatial features represent information from individual views, and global features learn the overall embedding of an LF image. We introduce the global feature extractor of the light field to enhance the performance of the LFSR model.

The MFSR network is composed of Feature Extract Blocks (FE Blocks) which are shown in Figure 2b. The input to the FE Block is a macro-pixel image feature, and the FE Block uses five parallel branches: an angular branch, a spatial branch, two EPI branches, and a global branch to extract multiple features. The Angular Feature Extractor (AFE) and Spatial Feature Extractor (SFE) are just like LF-InterNet [32]. The EPI feature extractor (i.e., EPI_h and EPI_v) extracts horizontal and vertical EPI feature separately using 1D dilated convolutions. The Global Feature Extractor (GFE) using 3D convolutions extract overall LF features from stacking SAI. The feature fusion block concatenates multiple features and these features are fused by 2D convolutions. The features were obtained from the four feature extraction groups iteratively, then the feature was transformed into sub-aperture form and used to calculate the SR local residual results. Finally, the low-resolution SAI array Bilinear Up-sampling [34] element-wise added with the residual results to output the SR results.



(a) Overall Architecture

(c) Global Feature Extract(GFE)

(b) Feature Extract Block(FE Block)

Figure 2. (**a**) The architectural overview of the MFSR. (**b**) Feature Extract Block(FEB). (**c**) Global Feature Extract Block(GFE). All up-sampling operator is the Bilinear Upsampling.

To extract the spatial features of the 4D light field, which are the features of a continuous region in the different sub-apertures of the light field, the macro-pixel images are reflected as pixels with the same angular coordinates in images from different sub-apertures. Just like LF-InterNet [32], we apply dilation convolution to treat the pixels at corresponding positions. For efficient extraction of light field spatial features, the spatial feature extractor as a convolution whose convolution kernel size is 3 × 3 with a stride of 1 and dilation as A (angular resolution) and executes zero padding to assure the same size of spatial resolution of input and output macro-pixel images.

For angular feature extraction in the 4D light field, which is pixels at the same angular coordinate from different views, the corresponding pixels to different angles could be properly organized for the macro-pixel images as a macro-pixel represented in them. Just like LF-InterNet [32], the angular feature extractor in our MFSR is conducted by 2D convolution, where the size of the convolution kernel is $A \times A$ whose stride is 1 and takes 0 as padding in the horizontal direction. The utilization of the convolution kernel with the size of $A \times A$ could exactly assure pixels in a macro-pixel are extracted by corresponding features.

Spatial and angular features are unable to efficiently represent the relation between angular and spatial for light field feature representation. The EPI image is always applied to represent the relation features of angular and spatial in the dual-view or multi-view images. The research thus applies a 1D convolution kernel for EPI feature extraction employing the light field as a multi-view image. The horizontal feature extractor applies 1D convolution which is to utilize 1D dilated convolution, the kernel size set to $1 \times A^2$, and the stride is $(1, A)$. Furthermore, no padding is performed to ensure that the size of the output feature space is consistent with the input feature image, where $A \times (A-1)/2$ is the angular resolution of the 4D LF. The EPI feature extraction is similar to that in the horizontal direction.

We propose a global feature extractor to learn the inner features of 4D LF. Although the relations of different directions are separately extracted in the above feature extraction blocks, the EPI features in the two directions could only represent the features in the two directions and are unable to sufficiently extract the relationships between different views. Therefore, we apply 3D convolution to conduct the extraction of the LF images global features which applies 3 × 3 × 3 convolution kernel whose stride is 1 to ensure the size of output feature images is the same as the input. This convolution method could efficiently obtain global features and learn the inner features of the needy implicit representation.

We design a multiple-feature fusion block to integrate spatial, angular, EPI, and global feature. At first, the resolutions of the above features are aligned by pixelShuffle1D [11] and 1 × 1 convolution. Then, all the features are attached to the channel dimension. At last, the multi-dimension of features is converted by 1 × 1 convolution to the target dimension. Taking the resolution of spatial features as the base for aligning other features, the angular features are converted into spatial feature resolution by pixelShuffle2D [11]. The resolution conversion of EPI features in the two directions is conducted by pixelShuffle1D [11]. The global features could be transformed to corresponding spatial feature resolution after 3D convolution.

To realize more sufficient feature extraction and integration, inspired by LF-InterNet [32], we cascade several feature extraction blocks iteratively to extract multiple features of LF image. The extracted features are sent to the up-sampling block for spatial up-sampling to obtain light field image features with target resolution. Furthermore, the up-sampling operation increases the number of the channels to $\alpha^2 \times C$ (α is the super-resolution scale, C is the feature channels) by 1 × 1 convolution. Then, the up-sampling of the feature images is conducted through pixelShuffle2D [11]. The ultimate results are reduced dimension through 1 × 1 convolution single channel, which could obtain the Y channel images of ultimate super-resolution by adding the up-sampling results in the original images.

The channel attention mechanism [29] is used to improve the integration of extracted angular and spatial features in the proposed method. This mechanism allows the network to pay more attention to important regional features, such as high-frequency and marginal feature information in the light field. By combining the attention mechanism with global feature integration, the method can effectively drive the integration of light field super-resolution features, leading to better performance in terms of PSNR and SSIM scores, as demonstrated in the experiments.

3.2.2. Loss Function

Inspired by DPT [13], the gradient images are able to assist the super-resolution model to learn detailed texture information. Therefore, the gradient loss function is used for rich texture with the LF image efficiently. The loss function of MFSR can be represented by Equation (1). I and $\hat{I}$, respectively, represent the super-resolved image and the ground truth image, (i, j) is the pixel location, (M, N) is the image resolution, and $Sobel(I)$ represent a gradient detector [14] for image I.

$$Loss = \alpha_1 L_1(I, \hat{I}) + \alpha_2 L_{gradient}(I, \hat{I}) \tag{1}$$

$$L_1(I, \hat{I}) = \frac{1}{N}\frac{1}{M}\sum_{i=1}^{N}\sum_{j=1}^{M}|I(i,j) - \hat{I}(i,j)| \tag{2}$$

$$L_{gradient}(I, \hat{I}) = L_1(Sobel(I), Sobel(\hat{I})) \tag{3}$$

In the gradient loss, the Sobel operator [14] is a gradient extraction function that separately extracts the horizontal and vertical gradients through two convolutions. Through the gradient extractor, the results of bicycle and origami scenes are visualized as shown in Figure 3. It can be seen that the gradient information in the image can be effectively extracted to assist the learning of the image super-resolution network.

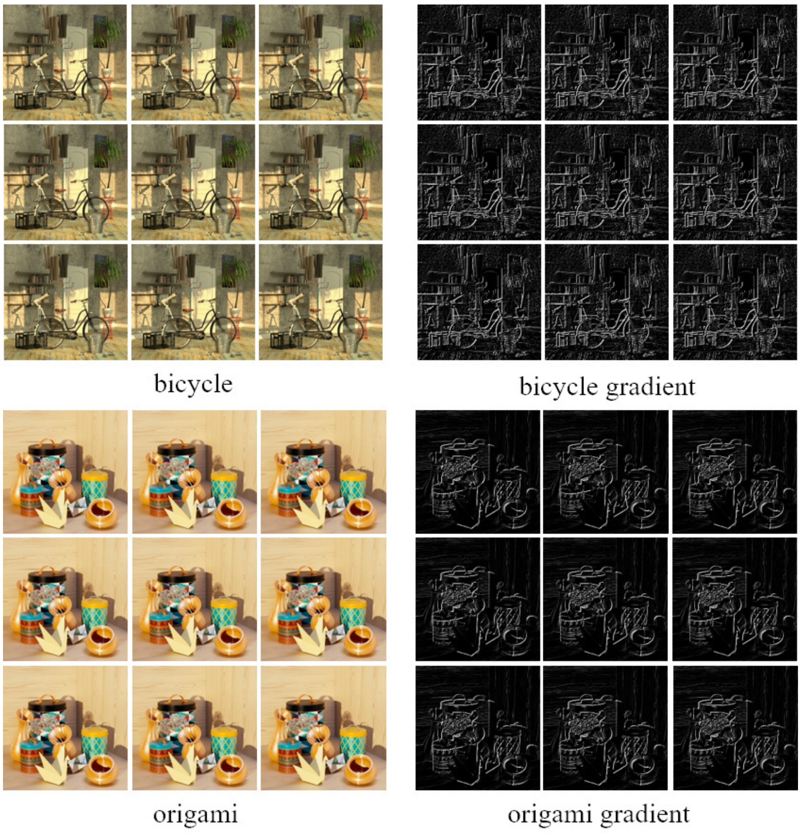


Figure 3. Original and Bicycle scene image (**left**) and their gradient image using Sobel Operator [14].

4. Experiments

In this section, we first present the datasets and implementation details. Then, we demonstrate the effectiveness of our proposed MFSR through an ablation study. We conduct a discussion and analysis of our algorithm based on the experiments. Finally, we compare our proposed MFSR to several state-of-the-art light field image super-resolution methods. The training setup and ablation study setup as shown in Figure 4.

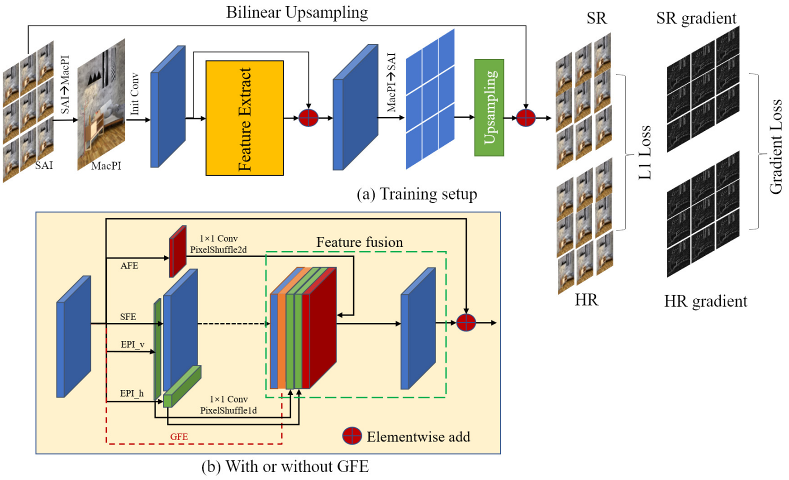


Figure 4. Experiments setup. (**a**) Training model by Adam optimizer [35] and two losses. (**b**) With/without Global feature extract for MFSR.

4.1. Datasets and Implementation Details

We conduct extensive experiments on 5 open light field images datasets including EPFL [16], HCI_new [15], HCI_old [36], INRIA_Lytro [37], and STF_gantry [38]. The angular resolution of the whole light field images in the above datasets is all 9 × 9 $(U = V = A = 9)$, and PSNR and SSIM are ultimate evaluation indicators of the super-resolution model. For test datasets with multiple scenarios, the researchers obtain the sub-aperture image calculation evaluation indicator at a certain angular resolution first and average the evaluation indicator results of different sub-apertures in multiple scenarios, and obtain the evaluation indicator value of this dataset last. The angular resolution in the experiments of this paper is $U = V = A = 5$ and the selected test dataset scenarios division in the experiment is as shown in Table 1.

Table 1. Datasets used in our experiments.

Dataset	HCI_NEW	Stanford_Gantry	EPFL	INRIA_Lytro	HCI_OLD	Total
Train	20	9	70	35	10	144
Test	4	2	10	5	2	23

For model training, the image space is converted from RGB to YCbCr in this paper, and the model only inputs the images of the Y channel for training. During the test, the PSNR and SSIM indicators calculated from the Y channel of the light field image are directly used for model performance evaluation. In the period of training, each sub-aperture is cut into blocks with a stride of 32, and the spatial resolution of each sample is 160 × 160 by setting the angular resolution of the light field as $U = V = 5$. After down-sampling, the spatial resolution of the sample is 80 × 80. At the same time, this research also conducts the number enhancement of random horizontal flip, vertical flip, and 90° rotation in the model training process to improve the generalization ability of the model. In the process of data enhancement, the researchers transform the spatial dimension and angular dimension simultaneously to maintain the structural consistency of the light field before and after data enhancement.

The model in this paper is trained with the loss function of Equation (1), set loss weight $\alpha_1 = \alpha_2 = 1$, and optimized by Adam [35] optimizer, $\beta_1 = 0.9$, $\beta_2 = 0.999$, and the batch size is set to 4. All experiments in this study are conducted based on PyTorch, and the model is optimized by the Adam [35] optimizer. A single NVidia V100 GPU was used for model training. The initial learning rate is set to 2×10^{-4}, and the learning rate is reduced by a factor of 0.5 every 15 epochs. The training is stopped after 50 epochs, and the weights of the 50th epoch are finally adopted as the weight for model examination.

4.2. Ablation Study

To verify the validation of our proposed MFSR, we compared for performance influence on the light field image super-resolution from global feature and gradient loss in this section. In the DistgSSR [11], the influence on the performance of the light field images super-resolution caused by spatial, angular, and EPI features which finally demonstrates the effectiveness of sub-space features (spatial, angular, and EPI features).

To verify the validation of global features for light field image super-resolution, we introduce global feature fusion based on spatial, angular, and EPI features. The PSNR value comparison of the experimental results is shown in Table 2. It could be observed that PSNR values of the light field image super-resolution with the introduction of the global features in different datasets are higher than those without the introduction of the global features. Compared to the original model, the PSNR value of 4 times super-resolution with the introduction of the global features is promoted by 0.219 dB. That indicates the superiority of integration with global features of the light field images compared to mere utilization of local features which efficiently promotes the performance of the light field image super-resolution.

Table 2. Ablation study of Global Feature in the proposed model. We compare the performance of the model with (w) or without (w/o) Global Feature on different test datasets and report the PSNR results.

Global Feature	HCI_new	Stanford_Gantry	EPFL	INRIA_Lytro	HCI_old	Average
w	31.11	30.891	28.711	30.836	37.198	31.749
w/o	31.239	31.223	28.943	30.959	37.48	31.968

To demonstrate the influence on the light field image spatial super-resolution of the loss function introduced by us, the paper introduces a gradient loss function to train the model while introducing the global feature. Furthermore, the experiment results in comparison are shown in Table 3. It is obvious that the average PSNR after the introduction of the gradient loss function is promoted by 0.1 dB but the PSNR of the model declines in the dataset EPFL and HCI_old. The researchers conclude that the corresponding light field scenario gradient information is not evident enough in these datasets. The majority of these scenarios are relatively smooth making the gradient information have little assistance to super-resolution in these scenarios.

Table 3. Ablation study of Gradient Loss in the proposed model. We compare the performance of the model with (w) or without (w/o) Gradient Loss on different test datasets and report the PSNR results.

Gradient Loss	HCI_NEW	Stanford_Gantry	EPFL	INRIA_Lytro	HCI_OLD	Average
w	31.239	31.223	28.943	30.959	37.48	31.968
w/o	31.278	31.338	28.92	30.966	37.388	31.978

4.3. Experimental Results on the Light Field Dataset

The light field images spatial super-resolution algorithm put forward by us is compared with several of the most advanced algorithms including an algorithm with a baseline of bicubic interpolation, 2 traditional image super-resolution algorithms (EDSR [28], RCAN [29]), 7 light field images super-resolution algorithms(resLF [10], LFATO [31], LF-InterNet [32], LF-DFNet [33], MEG-Net [9], DPT [13] and DistgSSR [11]). In this research, the light field super-resolution algorithms based on deep learning in the above methods are trained and examined in the same train datasets and test datasets for fair performance comparison of every method. Due to that model training takes a great deal of time, the paper only provides experiment results of 2× and 4× super-resolution with ab angular resolution of 5 × 5 and PSNR and SSIM of related experiments as shown in Table 4.

As shown in Table 3, our proposed model obtains the highest PSNR and SSIM points from 5 datasets in the tasks of 2× and 4× LFSR. Under 2× and 4× SR, PSNR of our MFSR is promoted by 0.208 dB and 0.274 dB compared to the baseline model DistgSSR [11]. From the perspective of the performance from different datasets, our MFSR is promoted remarkably in the Stanford_Gantry and HCI_old and the PSNR is promoted by 0.447 dB and 0.282 dB compared to the baseline model DistgSSR [11] on 4× SR.

To demonstrate the superiority of this algorithm, the visualization analysis of light field test scenarios is shown in Figure 5. After the amplification of the local region, it could be observed that super-resolution results obtained by direct up-sampling are blurry in the details in rich texture. Compared with other algorithms for LFSR, due to the application of spatial, angular, EPI, and global features integration and utilization of the gradient loss function promoting the performance in the rich texture of super-resolution algorithm, the performance of our MFSR is better in the rich texture. Compared to the most advanced light field images spatial super-resolution algorithm currently, MFSR could produce more realistic images with fewer ghosts.

Table 4. PSNR/SSIM values achieved by different methods for 2× and 4× SR. The best results are in red and the second best results are in green.

Methods	Scale	HCI_new	Stanford_Granty	EPFL	INRIA_Lytro	HCI_old
Bicubic	×2	31.887/0.9356	31.063/0.9498	29.740/0.9376	31.331/0.9577	37.686/0.9785
EDSR [28]	×2	34.828/0.9592	36.296/0.9818	33.089/0.9629	34.985/0.9764	41.014/0.9874
RCAN [29]	×2	35.022/0.9603	36.670/0.9831	33.159/0.9634	35.046/0.9769	41.125/0.9875
resLF [10]	×2	36.685/0.9739	38.354/0.9904	33.617/0.9706	35.395/0.9804	43.422/0.9932
LF-ATO [31]	×2	37.244/0.9767	39.636/0.9929	34.272/0.9757	36.170/0.9842	44.205/0.9942
LF_InterNet [32]	×2	37.319/0.9772	38.838/0.9917	34.298/0.9762	36.108/0.9847	44.534/0.9945
LF-DFnet [33]	×2	37.418/0.9773	39.427/0.9926	34.513/0.9755	36.416/0.9840	44.198/0.9941
MEG-Net [9]	×2	37.424/0.9777	38.767/0.9915	34.312/0.9773	36.103/0.9849	44.097/0.9942
DPT [13]	×2	37.355/0.9771	39.429/0.9926	34.490/0.9758	36.409/0.9843	44.302/0.9943
DistgSSR [11]	×2	37.838/0.9791	40.341/0.9940	34.306/0.9773	36.247/0.9853	44.826/0.9948
MFSR	×2	37.964/0.9943	40.560/0.9943	34.859/0.9791	36.518/0.9861	44.699/0.9947
Bicubic	×4	27.715/0.8517	26.087/0.8452	25.264/0.8324	26.952/0.8867	32.576/0.9344
EDSR [28]	×4	29.591/0.8869	28.703/0.9072	27.833/0.8854	29.656/0.9257	35.176/0.9536
RCAN [29]	×4	29.694/0.8886	29.021/0.9131	27.907/0.8863	29.805/0.9276	35.359/0.9548
resLF [10]	×4	30.723/0.9107	30.191/0.9372	28.260/0.9035	30.338/0.9412	36.705/0.9682
LF-ATO [31]	×4	30.880/0.9135	30.607/0.9430	28.514/0.9115	30.711/0.9484	36.999/0.9699
LF_InterNet [32]	×4	30.998/0.9166	30.537/0.9432	28.737/0.9143	30.701/0.9485	37.101/0.9714
LF-DFnet [33]	×4	31.136/0.9177	31.035/0.9481	28.685/0.9141	30.770/0.9497	37.175/0.9711
MEG-Net [9]	×4	30.882/0.9146	30.437/0.9415	28.538/0.9107	30.542/0.9463	36.861/0.9696
DPT [13]	×4	31.028/0.9161	30.770/0.9451	28.604/0.9142	30.681/0.9486	37.098/0.9706
DistgSSR [11]	×4	31.110/0.9175	30.891/0.9466	28.711/0.9149	30.836/0.9492	37.198/0.9715
MFSR	×4	31.327/0.9207	31.262/0.9506	28.985/0.9183	31.024/0.9515	37.519/0.9729

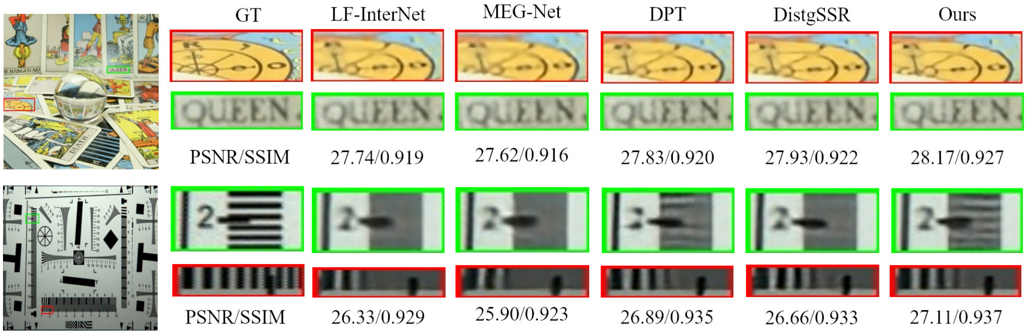


Figure 5. Comparison visualization of our MFSR and other LFSR methods.

5. Conclusions

We present an end-to-end LFSR model called MFSR that utilizes multiple features to integrate spatial, angular, EPI, and global features. The MFSR extracts multiple features from the LF using a feature convolutional neural network. To promote the performance of LFSR, the global features are extracted from the sub-aperture images through 3D convolution. Furthermore, to preserve detailed texture information that may be lost during the LFSR process, we apply a gradient loss based on Sobel operator [14]. Experimental results show that our algorithm achieves the highest PSNR and SSIM scores in the task of 2× and 4× super-resolution. Even in complex textures, our algorithm can produce high-quality super-resolution images. We also conduct ablation experiments to demonstrate the effectiveness of our proposed MFSR.

Additionally, future research could also focus on exploring the possibility of integrating other imaging modalities with the light field, such as hyperspectral or polarimetric information. Moreover, the application of reinforcement learning or generative adversarial networks (GANs) could be explored to enhance the performance of LFSR. Lastly, the development of a more extensive benchmark dataset for LFSR could facilitate the comparison and evaluation of different algorithms, and promote the advancement of t he field.

Author Contributions: Conceptualization, J.Z. and H.W.; Data curation, J.Z. and H.W.; Formal analysis, J.Z. and H.W.; Investigation, J.Z. and H.W.; Methodology, J.Z. and H.W.; Project administration, H.W.; Resources, Honintegrate; Software, J.Z.; Supervision, J.Z.; Validation, J.Z.; Visualization, J.Z.; Writing—the original draft, J.Z.; Writing—review & editing, J.Z. and H.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The public light field datasets used in this study are EPFL [16], HCI_new [15], HCI_old [36], INRIA_Lytro [37], and STF_gantry [38]. Details about these Light Field Datasets can be found in the relevant references. Training and testing dataset split using LFSR benchmark https://github.com/ZhengyuLiang24/BasicLFSR (accessed on 10 February 2023).

Acknowledgments: Thanks to Zhengyu Liang for establishing the light field super-resolution benchmark https://github.com/ZhengyuLiang24/BasicLFSR (accessed on 10 February 2023).

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Wang, Y.; Yang, J.; Guo, Y.; Xiao, C.; An, W. Selective light field refocusing for camera arrays using bokeh rendering and superresolution. *IEEE Signal Process. Lett.* **2018**, *26*, 204–208. [CrossRef]
2. Bishop, T.E.; Favaro, P. The light field camera: Extended depth of field, aliasing, and superresolution. *IEEE Trans. Pattern Anal. Mach. Intell.* **2011**, *34*, 972–986. [CrossRef] [PubMed]
3. Wang, T.-C.; Efros, A.A.; Ramamoorthi, R. Depth estimation with occlusion modeling using light-field cameras. *IEEE Trans. Pattern Anal. Mach. Intell.* **2016**, *38*, 2170–2181. [CrossRef] [PubMed]
4. Park, I.K.; Lee, K.M. Robust light field depth estimation using occlusion-noise aware data costs. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *40*, 2484–2497.
5. Zhou, M.; Ding, Y.; Ji, Y.; Young, S.S.; Yu, J.; Ye, J. Shape and reflectance reconstruction using concentric multi-spectral light field. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *42*, 1594–1605. [CrossRef] [PubMed]
6. Alperovich, A.; Johannsen, O.; Strecke, M.; Goldluecke, B. Light field intrinsics with a deep encoder-decoder network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 9145–9154.
7. Wang, Y.; Wu, T.; Yang, J.; Wang, L.; An, W.; Guo, Y. Deoccnet: Learning to see through foreground occlusions in light fields. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Snowmass Village, CO, USA, 1–5 March 2020; pp. 118–127.
8. Li, Y.; Yang, W.; Xu, Z.; Chen, Z.; Shi, Z.; Zhang, Y.; Huang, L. Mask4d: 4d convolution network for light field occlusion removal. In Proceedings of the ICASSP 2021–2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Toronto, ON, Canada, 6–11 June 2021; pp. 2480–2484.
9. Zhang, S.; Chang, S.; Lin, Y. End-to-end light field spatial super-resolution network using multiple epipolar geometry. *IEEE Trans. Image Process.* **2021**, *30*, 5956–5968. [CrossRef] [PubMed]
10. Zhang, S.; Lin, Y.; Sheng, H. Residual networks for light field image super-resolution. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 16–20 June 2019; pp. 11046–11055.
11. Wang, Y.; Wang, L.; Wu, G.; Yang, J.; An, W.; Yu, J.; Guo, Y. Disentangling light fields for super-resolution and disparity estimation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2022**, *45*, 425–443. [CrossRef] [PubMed]
12. Jiang, L.; Dai, B.; Wu, W.; Loy, C.C. Focal frequency loss for image reconstruction and synthesis. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 11–17 October 2021; pp. 13919–13929.
13. Wang, S.; Zhou, T.; Lu, Y.; Di, H. Detail-preserving transformer for light field image super-resolution. In Proceedings of the AAAI Conference on Artificial Intelligence, Online, 22 February–1 March 2022; pp. 2522–2530.
14. Kittler, J. On the accuracy of the sobel edge detector. *Image Vis. Comput.* **1983**, *1*, 37–42. [CrossRef]
15. Honauer, K.; Johannsen, O.; Kondermann, D.; Goldluecke, B. A dataset and evaluation methodology for depth estimation on 4d light fields. In Proceedings of the Computer Vision–ACCV 2016: 13th Asian Conference on Computer Vision, Taipei, Taiwan, 20–24 November 2016; Springer: Berlin/Heidelberg, Germany, 2017; pp. 19–34.
16. Rerabek, M.; Ebrahimi, T. New light field image dataset. In Proceedings of the 8th International Conference on Quality of Multimedia Experience (QoMEX), Lisbon, Portugal, 6–8 June 2016.
17. Liang, C.-K.; Ramamoorthi, R. A light transport framework for lenslet light field cameras. *ACM Trans. Graph. (TOG)* **2015**, *34*, 1–19. [CrossRef]

18. Alain, M.;Smolic, A. Light field super-resolution via lfbm5d sparse coding. In Proceedings of the 2018 25th IEEE international conference on image processing (ICIP), Athens, Greece, 7–10 October 2018; pp. 2501–2505.
19. Wanner, S.; Goldluecke, B. Variational light field analysis for disparity estimation and super-resolution. *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**, *36*, 606–619.
20. 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 the 2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, Washington, DC, USA, 16–21 June 2012; pp. 22–28.
21. Bishop, T.E.; Zanetti, S.; Favaro, P. Light field superresolution. In Proceedings of the 2009 IEEE International Conference on Computational Photography (ICCP), San Francisco, CA, USA, 16–17 April 2009; pp. 1–9.
22. Cho, D.; Lee, M.; Kim, S.; Tai, Y.W. Modeling the calibration pipeline of the lytro camera for high quality light-field image reconstruction. In Proceedings of the IEEE International Conference on Computer Vision, Sydney, Australia, 1–8 December 2013 pp. 3280–3287.
23. Rossi, M.; Frossard, P. Geometry-consistent light field super-resolution via graph-based regularization. *IEEE Trans. Image Process.* **2018**, *27*, 4207–4218.
24. Farrugia, R.A.; Galea, C.; Guillemot, C. Super resolution of light field images using linear subspace projection of patch-volumes. *IEEE J. Sel. Top. Signal Process.* **2017**, *11*, 1058–1071.
25. Dong, C.; Loy, C.C.; He, K.; Tang, X. Learning a deep convolutional network for image super-resolution. In Proceedings of the Computer Vision—ECCV 2014: 13th European Conference, Zurich, Switzerland, 6–12 September 2014; Springer: Berlin/Heidelberg, Germany, 2014; pp. 184–199.
26. Ledig, C.; Theis, L.; Huszár, F.; Caballero, J.; Cunningham, A.; Acosta, A.; Aitken, A.; Tejani, A.; Totz, J.; Wang, Z.; et al. Photo-realistic single image super-resolution using a generative adversarial network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 4681–4690.
27. Kim, J.; Lee, J.K.; Lee, K.M. Accurate image super-resolution using very deep convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 21–23 June 1994; pp. 1646–1654.
28. Lim, B.; Son, S.; Kim, H.; Nah, S.; Mu Lee, K. Enhanced deep residual networks for single image super-resolution. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Honolulu, HI, USA, 21–26 July 2017; pp. 136–144.
29. Zhang, Y.; Li, K.; Li, K.; Wang, L.; Zhong, B.; Fu, Y. Image super-resolution using very deep residual channel attention networks. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 286–301.
30. Yoon, Y.; Jeon, H.G.; Yoo, D.; Lee, J.Y.; So Kweon, I. Learning a deep convolutional network for light-field image super-resolution. In Proceedings of the IEEE International Conference on Computer Vision Workshops, Santiago, Chile, 7–13 December 2015; pp. 24–32.
31. Jin, J.; Hou, J.; Chen, J.; Kwong, S. 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, Seattle, WA, USA, 13–19 June 2020; pp. 2260–2269.
32. Wang, Y.; Wang, L.; Yang, J.; An, W.; Yu, J.; Guo, Y. Spatial-angular interaction for light field image super-resolution. In Proceedings of the Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, 23–28 August 2020; Springer: Berlin/Heidelberg, Germany, 2020; pp. 290–308.
33. Wang, Y.; Yang, J.; Wang, L.; Ying, X.; Wu, T.; An, W.; Guo, Y. Light field image super-resolution using deformable convolution. *IEEE Trans. Image Process.* **2020**, *30*, 1057–1071.
34. Youssef, A. Image downsampling and upsampling methods. *Natl. Inst. Stand. Technol.* **1999**. *10*, 57–61
35. Niklaus, S.; Mai, L.; Liu, F. Video frame interpolation via adaptive separable convolution. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 261–270.
36. Wanner, S.; Meister, S.; Goldluecke, B. Datasets and benchmarks for densely sampled 4d light fields.; In Proceedings of the VMV, Bayreuth, Germany, 10–12 October 2013; pp. 225–226.
37. Le Pendu, M.; Jiang, X.; Guillemot, C. Light field inpainting propagation via low rank matrix completion. *IEEE Trans. Image Process.* **2018**, 27, 1981–1993. [CrossRef] [PubMed]
38. Vaish V.; Adams A. The (new) stanford light field archive. In *Computer Graphics Laboratory*; Stanford University: Stanford, CA, USA, 2008; Volume 6, p. 7.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.