

# Reivew of Light Field Image Super-Resolution

Li Yu <sup>1,\*</sup>, Yunpeng Ma<sup>1</sup>, Song Hong<sup>2</sup> and Ke Chen<sup>3</sup>

<sup>1</sup> Engineering Research Center of Digital Forensics, Ministry of Education, Nanjing University of Information Science and Technology, Nanjing 210044, China; 20211249438@nuist.edu.cn

<sup>2</sup> Air Ammunition Research Institute Co., Ltd., Norinco Group, Beijing 100053, China; abc441759927@126.com

<sup>3</sup> School of Electronic and Information Engineering, South China University of Technology, Guangzhou 510641, China; chenks@scut.edu.cn

\* Correspondence: li.yu@nuist.edu.cn

**Abstract:** Currently, light fields play important roles in industry, including in 3D mapping, virtual reality and other fields. However, as a kind of high-latitude data, light field images are difficult to acquire and store. Thus, the study of light field super-resolution is of great importance. Compared with traditional 2D planar images, 4D light field images contain information from different angles in the scene, and thus the super-resolution of light field images needs to be performed not only in the spatial domain but also in the angular domain. In the early days of light field super-resolution research, many solutions for 2D image super-resolution, such as Gaussian models and sparse representations, were also used in light field super-resolution. With the development of deep learning, light field image super-resolution solutions based on deep-learning techniques are becoming increasingly common and are gradually replacing traditional methods. In this paper, the current research on super-resolution light field images, including traditional methods and deep-learning-based methods, are outlined and discussed separately. This paper also lists publicly available datasets and compares the performance of various methods on these datasets as well as analyses the importance of light field super-resolution research and its future development.

**Keywords:** light field; image super-resolution; deep learning; convolutional neural networks



**Citation:** Yu, L.; Ma, Y.; Hong, S.; Chen, K. Reivew of Light Field Image Super-Resolution. *Electronics* **2022**, *11*, 1904. <https://doi.org/10.3390/electronics11121904>

Academic Editor: Nikolay Hinov

Received: 9 May 2022

Accepted: 14 June 2022

Published: 17 June 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



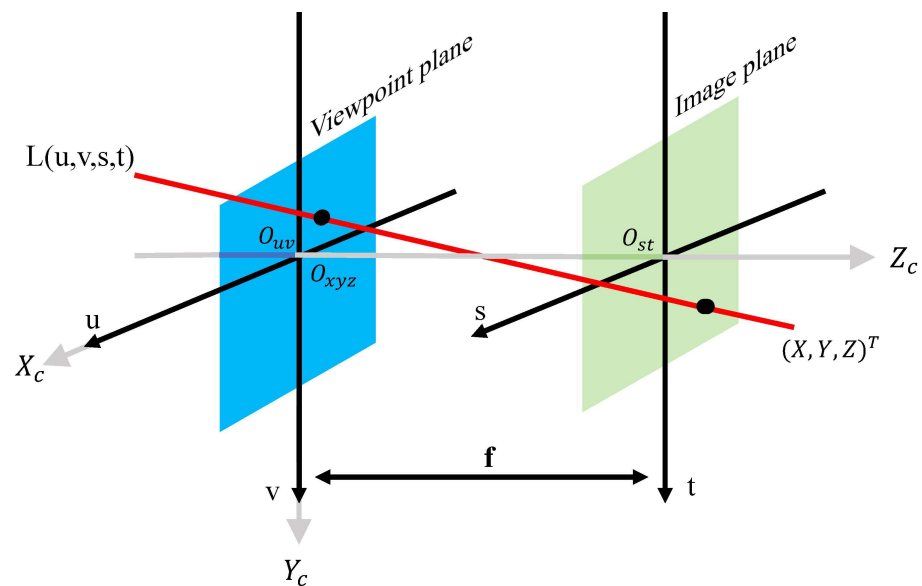
**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The eye can see objects in the world because it receives the light emitted or reflected by the object. The light field is a complete representation of the collection of light in the three-dimensional world. Therefore, collecting and displaying the light field can visually reproduce the real world to a certain extent. In 1846, Michael Faraday [1] proposed the idea of interpreting light as a field.

Gershun [2] introduced the concept of a “light field” in 1936 by representing the radiation of light in space as a three-dimensional vector of spatial positions. Adelson and Bergen [3] further refined the work of Gershun [2] in 1991, and they proposed the “Plenoptic Function”, which uses five dimensions to represent light in the three-dimensional world. Levoy [4] reduced the 5-dimensional Plenoptic function to four dimensions by fixing the intensity of the light during propagation, which is now called a 4D light field.

As shown in Figure 1, the model proposed by Levoy uses two planes to simultaneously record the angle and position information of light in the space.  $L(u, v, s, t)$  represents a sample of light field, where  $L$  represents the light intensity. The viewpoint plane  $(u, v)$  is located on the  $Z = 0$  plane and records the direction information of the light. The image plane  $(s, t)$  is parallel to the viewpoint plane and is located on the plane of the camera coordinate system  $Z = f$ , which records the position information of the light ( $f$  as the distance between the two planes). Any ray emitted from a point  $(X, Y, Z)$  in space can be uniquely determined by knowing its intersection with the viewpoint plane  $(u, v)$  and the image plane  $(s, t)$ .



**Figure 1.** The two-plane parametric representation of the four-dimensional light field.

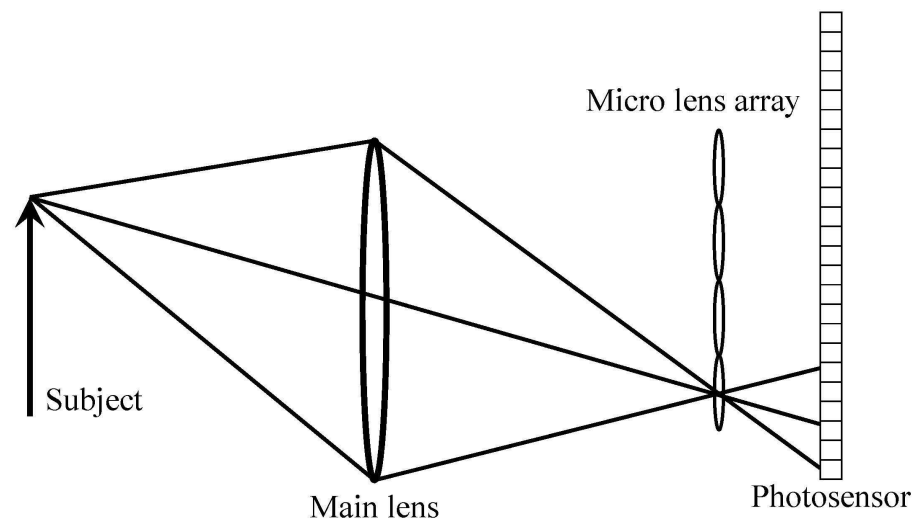
As a kind of high-dimensional data, light field data is difficult to be formally expressed in the three-dimensional world. Therefore, the early collection of 4D light field images requires special light field cameras [5,6]. As shown in Figure 2, the light field camera embeds a micro lens array between the main lens of the traditional camera and the photosensor. Light going through the main lens will be projected onto the photosensor plane after passing through the micro lens units on the micro lens array to form a unit image.

If each unit image is regarded as a macro-pixel, the points at the same position of each macro-pixel correspond to samples of the same direction at different positions of the photographed object. The image array, generated by extracting the pixels at the same position in each macro-pixel, can form a sub-image array of different directions, i.e., sub-aperture image, which contains both angular and spatial information of the photographed object. These pixels can form a sub-aperture image together. These sub-aperture images are the images formed by each angle of the light field so that angle information and spatial information can be captured at the same time.

With the continuous deepening of research, a variety of light field image acquisition methods have been developed. For example, multi-camera array imaging [7,8] uses multiple cameras at different spatial positions to collect images of different perspectives and then reconstruct the 4D light field data of the shooting scene. The spatial resolution of the light field data acquired by the camera array is determined by the sensor size of a single camera, and the angular resolution is determined by the number of cameras.

By arranging multiple cameras into different arrays and adjusting the distance between the cameras or the imaging plane, different imaging effects can be achieved; the method of an encoding mask [9,10] is to insert a programmable mask between the main lens and the sensor to lightly modulate the light entering the camera aperture and then reconstruct the light field through linear or non-linear algorithms. This method can learn to remove the redundancy of the light field and collect less data to reconstruct the complete light field.

Although there are various methods of light field collection, the light field images collected by these methods have various problems. For example, although the micro lens array can form a light field image through a single shot, its spatial resolution and angular resolution are inadequate for generating a clear image; while the light field data acquired by encoding masks can improve the angular resolution without sacrificing the image resolution, the peak signal-to-noise ratio (PSNR) of its acquired data is low.



**Figure 2.** Schematic diagram of light field camera imaging.

In order to optimize the collected light field image, it is necessary to perform super-resolution processing [11]. Early light field super-resolution methods mainly include geometric projection [12] and optimization using prior knowledge. The projection is mainly based on the imaging principle of the light field camera, using rich sub-pixel information to propagate the pixels of each sub-aperture image to the target view for super-resolution. Nava [13] obtained inspiration from the Focal Stack transformation and developed a projection-based technology. The method based on optimization mainly relies on different mathematical models to perform super-resolution processing on a light field under various optimization frameworks. Bishop [14] performed this task by means of a variational Bayesian framework.

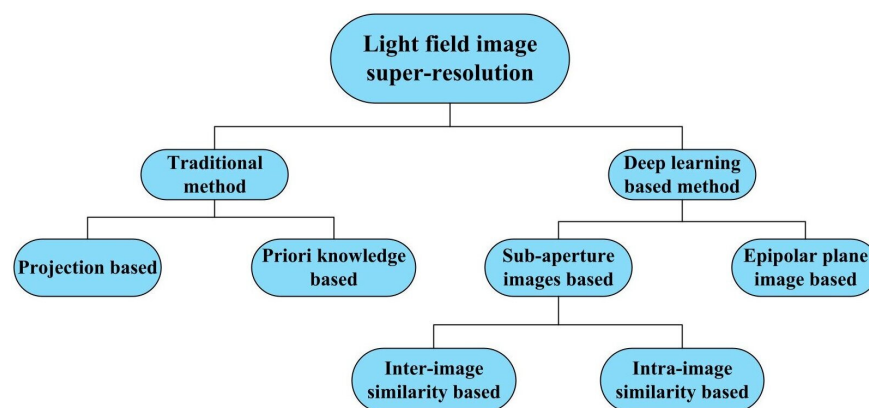
With the boom in artificial intelligence in recent decades, deep learning has proven its effectiveness in many fields, including image super-resolution [15], image depth estimation [16], object detection [17], face recognition [18–20] and biometrics [21]. At the same time, deep learning is also used in the task of light field super-resolution. The method proposed by Yoon [22] laid a solid foundation for the combination of deep learning and light field super-resolution.

Currently, there are many kinds of neural networks that can be used for super-resolution of light fields. For example, Wang [23] used circular convolution to improve the spatial resolution of light fields; Zhang [24] used residual convolutional networks to reconstruct light field images with high spatial resolution. It is feasible to perform light field super-resolution through neural networks, and the development prospects are quite broad.

Benefiting from the rapid development of virtual reality technology, light field technology has received increasing attention. Some light-field-related reviews have been published in recent years. Guillemot [25] and Ihrke [26] gave a general review of the light-field-related research, including the light field principle and light-field-rendering-related contents, while not involving light field super-resolution. Wu [27] provided a comprehensive review of light-field-related research, where light field image super-resolution is also included.

However, many new deep-learning-based light field super-resolution methods have been proposed since 2017, which were not included in [27]. Thus, an up-to-date review of light field super-resolution methods is needed. In this review, papers related to light field super-resolution since 2009 were searched in databases, including IEEE, Springer, and Elsevier. Keywords, such as “light field super-resolution”, “light field reconstruction”, “sub-aperture images”, “epipolar plane image (EPI)” and “deep learning” were used. Specifically, the main contributions of this paper are as follows:

1. We present a comprehensive review of light field image super-resolution techniques, including problem settings, available datasets, performance metrics, and existing research methods.
2. We classify and illustrate light field super-resolution techniques in a hierarchical and structured manner, and summarize the possible factors affecting the quality of light field super-resolution through performance comparisons. The classification used in this review is shown in Figure 3 below.
3. We discuss the existing challenges in light field super-resolution processing tasks, and point out the possible future directions of light field super-resolution to provide a reference for other researchers.



**Figure 3.** Classification of light field image super-resolution methods in this review.

The structure of our article is as follows: Section 2 introduces the traditional light field super-resolution methods; Section 3 introduces the deep-learning-based light field super-resolution methods; Section 4 introduces the datasets and provides a comparative analysis of light field super-resolution methods; Section 5 indicates the challenges in the current light field super-resolution processing tasks and possible future development directions. Section 6 gives our conclusions. The abbreviations used in this review are summarized in Table 1.

**Table 1.** Full names of abbreviations.

Abbreviation	Full Name
LF	Light Field
EPI	Epipolar plane image
HR	High Resolution
LR	Low Resolution
LFCNN	Light Field Convolutional Neural Network
LFSR	Light Field Super-Resolution
SISR	Single Image Super-Resolution
VDSR	Very Deep Super-Resolution

## 2. Traditional Method

Super-resolution of light field images is the process of reconstructing a high-resolution light field image from a given low-resolution light field image. This section will mainly introduce the traditional super-resolution methods of light field images. What these traditional methods have in common is that there are no neural networks, no iterative training is required, and they rely solely on mathematical calculations and derivations to produce the desired results. These methods can be classified into two main categories: projection-based and priori-knowledge-based methods. Among them, priori-knowledge-based methods introduce prior knowledge from external sources.

### 2.1. Projection-Based LFSR

As introduced in the first section, the spatial resolution of the sub-aperture image is limited by the microlens resolution. The geometric projection-based approach calculates sub-pixel offsets between sub-aperture images of different views, based on which pixels in adjacent views can be propagated to the target view for super-resolution processing of the target view. Lim [28] indicated that the angular data in the 2D dimension of the light field contains information about the sub-pixel offsets of images in the spatial dimension from different viewpoints.

After extracting this information, the light field image can be super-resolution processed by projection onto convex sets (POCS) [29]. Nava [13] proposed a new super-resolution focus stack based on Fourier slice photographic transformation [30] and combined it with multi-view depth estimation to obtain super-resolution images. Pérez [31] extended the Fourier slicing technique to the super-resolution work of the light field and provided a new super-resolution algorithm based on Fourier slicing photography and discrete focus stack transform.

Yu [32] performed 2D integral projections of parametric light field samples. The analysis of the distribution of these projected samples in 2D space can yield a resolution enhancement factor for light field super-resolution processing. Zhou [33] extracted sub-pixel offsets between the angular image and the blur in the angular image, and with the help of this offset information, an observation model can be constructed from the high resolution image to the angular image. Wang [34] redefined the mapping function between the parallax and shear displacement of an image, and the proposed scheme based on this mapping function does not require additional a priori information. The advantage of the projection-based method is that it is simple and fast; however, this method has certain accuracy defects.

### 2.2. Priori-Knowledge Based LFSR

During the shooting process of the light field camera, due to the interference of external factors, such as the environment, light, and jitter, the obtained light field images often have low resolution and varying degrees of noise disturbance. In order to reconstruct a more realistic view with high resolution, a method based on a prior hypothesis was proposed. This type of method used the special high-dimensional structure of the 4D light field while adding priori assumptions about the actual shooting scene, and then proposed a mathematical model to optimize the solution of the super-resolution problem of the light field.

Bishop [14] incorporated Lambertian and texture preserving priors in a variational Bayesian framework to reconstruct scene depth and super-resolution textures, and the addition of scene depth information significantly improved the quality of the light field super-resolution. This algorithm has better performance results on real images. Mitra [35] used a Gaussian Mixture Model (GMM) to model the light field patch and found the disparity value of the patch through fast subspace projection technology, and then used the linear minimum mean-square error (LMMSE) algorithm to reconstruct the patch.

The method proposed by Rossi [36] used the different view information of the light field combined with graph regularization to enhance the light field structure and finally obtained a high-resolution view. Considering the noise problem in real light field images, Alain [37] proposed a method combining super-resolution block matching and 3D filtering (SRBM3D) [38] single image super-resolution filter and light field block matching and 5D filtering (LFBM5D) [39] light field noise reduction filter. The super-resolution of the light field is realized by repeatedly alternating the filtering steps and back-projection steps of LFBM5D [39]. Farrugia [40] used the estimated disparity information to reduce the matching area to improve the super-resolution quality.

Boominathan [41] used a low-resolution LF camera and a high-resolution digital single-lens reflex (DSLR) camera to form a hybrid imaging system and used a patch-based algorithm to combine the advantages of the two cameras to produce high-resolution images. The method

proposed by Pendu [42] based on the Fourier parallax layer model [43] can simultaneously solve various types of degradation problems in a single optimization framework.

### 3. Deep-Learning-Based Method

The prosperous development of deep learning has promoted the development of image super-resolution. The super-resolution convolutional neural network (SRCNN) proposed by Dong [44] in 2014 represented the end-to-end mapping between low/high resolution images. As shown in Figure 4 below, in order to learn this mapping, only three steps are required:

1. Patch extraction and representation: This operation extracts patches from low-resolution images and expresses them as high-dimensional vectors. The dimensionality of the vector is equal to the number of feature maps.
2. Non-linear mapping: This operation can non-linearly map the high-dimensional vector extracted in 1 to another high-dimensional vector, and each mapping vector can conceptually represent a high-resolution patch; these mapping vectors form another set of feature maps.
3. Reconstruction: This operation will operate the high-resolution patch set obtained in Step 2 to generate the final high-resolution image.

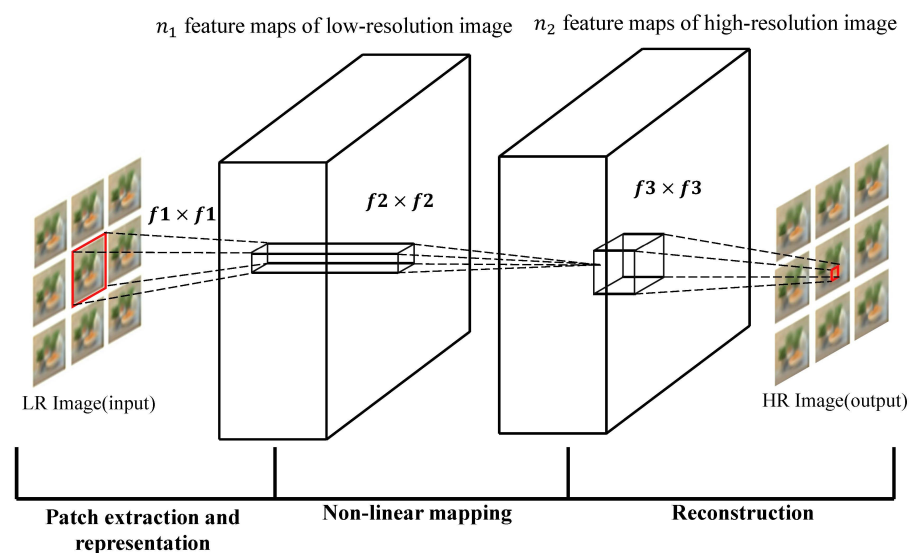


Figure 4. Pipeline of SRCNN [44].

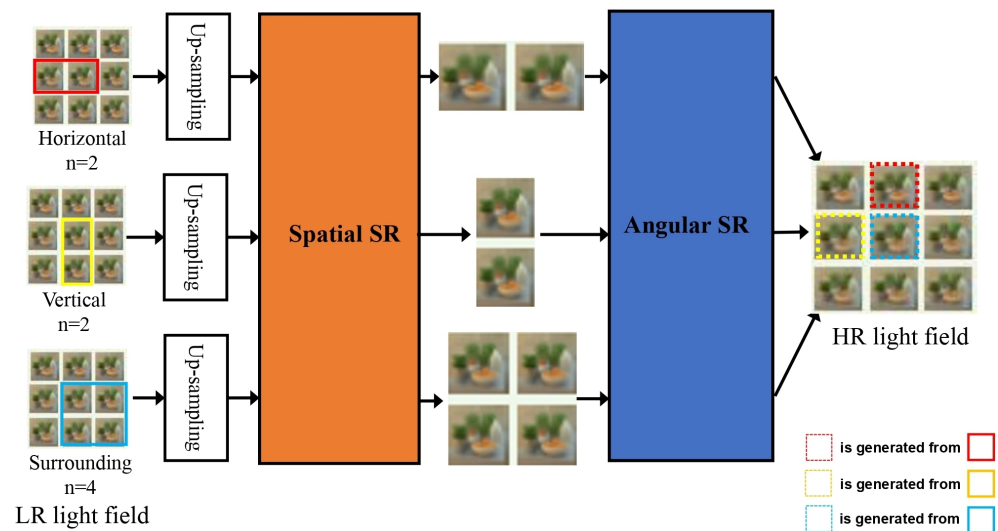
This kind of lightweight network structure achieved state-of-the-art recovery quality at that time, which was the first combination of deep learning and image super-resolution work. The subsequent network models for image super-resolution processing, such as very deep super-resolution network (VDSR) [45] and enhanced deep super-resolution network (EDSR) [46] were also inspired by it.

Although the good generalization ability of convolutional neural networks can provide enough training data to fit the model and cover a wide distribution of the expected test images, these super-resolution algorithms for single images cannot be directly applied to the super-resolution problem of light field images. Compared with the SISR work that only considers increasing the spatial resolution, the target of the light field super-resolution includes increasing both the angular resolution and the spatial resolution.

In 2015, Yoon [22] proposed a neural network model, which was named light field convolutional neural network (LFCNN), for light field image super-resolution, its overall structure is shown in Figure 5. The network model consists mainly of a spatial SR network and an angular SR network, with three different types of sub-aperture image pairs used as input throughout the network: horizontal pairs ( $n = 2$ ), vertical pairs ( $n = 2$ ) and surroundings ( $n = 4$ ). The spatial SR network is similar to [47] and can restore the high-

frequency details of the image. The angular SR network can generate new views between sub-aperture images, which is equivalent to increasing the number of sub-aperture images.

The special feature of LFCNN is that no matter how the depth and space of the scene change, the specific network layer used for angle and spatial resolution enhancement can restore the sub-aperture image well, thereby, improving the resolution of the image space domain and angular domain at the same time.



**Figure 5.** Overall framework of LFCNN [22].

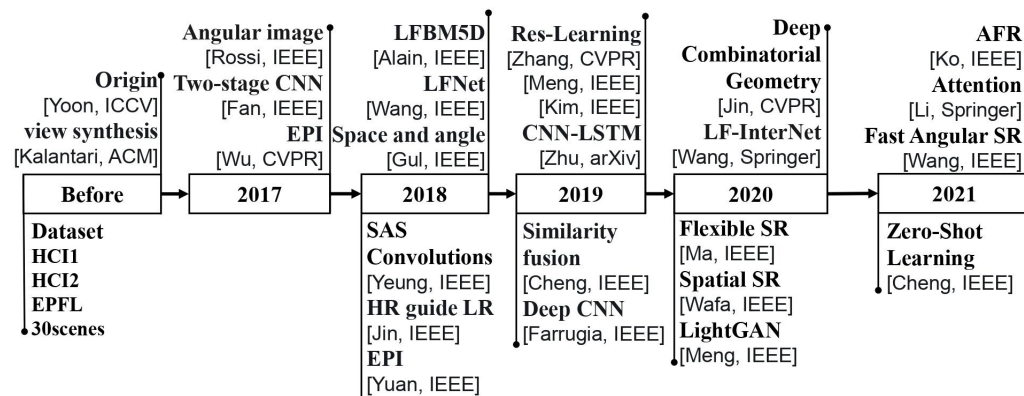
Although the method proposed by Yoon well combines deep learning and light field super-resolution and achieves good scores on both PSNR and structural similarity (SSIM); however, it also has shortcomings. In the design of the network structure, the three types of sub-aperture image pairs in the LFCNN are used as inputs into three separate CNN networks for single image super-resolution, which leads to different sub-aperture image pairs being processed independently of each other, ignoring the correlation information between them, and the complex network structure brings high computational costs and slows down the processing speed.

In terms of training data sets, Yoon used a Lytro Illum camera as the acquisition device to capture scenes with various textures, and a total of 300 photos were used for the training of the network, while the method proposed by Wanner [48] tested poorly on the dataset used by Yoon because it used a different dataset for training.

Figure 6 shows the development history of the light field super-resolution method based on deep learning. It can be seen from the figure that the study of light field super-resolution based on deep learning started in 2015, and many light field data sets have been proposed before, such as HCI [49,50], EPFL [51], etc. The main research objective of Rossi [36], Gul [52], Fan [53] and others during 2015–2018 was to further decouple the structure of the light field by convolutional neural networks and extract from it the angular and spatial information required for super-resolution processing.

Wu [27] and Yuan [54] et al. used EPI for super-resolution processing of the light field. With the prosperity of deep learning, many novel network structures have also been applied to the super-resolution processing of light field images, such as Zhang [24] used residual networks, Zhu [55] combined CNN with long short term memory (LSTM), Meng [56] used generative adversarial network (GAN). At present, the research on the super-resolution of light field images is not satisfied with higher super-resolution quality, however, has expanded to how to improve the processing speed of super-resolution while ensuring the super-resolution quality, such as Wang [57], Ma [58]. Few-shot learning has also been used in LFSR, where Cheng [59] proposed a zero-sample learning framework for LFSR.

The current work on deep-learning-based LFSR can be divided into two categories based on the input LF image type: sub-aperture images based and EPI based methods. The sub-aperture images based methods can be further classified into two types: intra-image similarity and inter-image similarity based methods.



**Figure 6.** Development timeline of light field super-resolution method based on deep learning.

### 3.1. Sub-Aperture-Image-Based LFSR

It is difficult to formally represent high-dimensional light field data in the three-dimensional world; however, fixing any two dimensions of the biplane light field model allows visualization of the light field by displaying a two-dimensional slice. Fixing the angular dimension enables the acquisition of light field data in the form of an array of sub-aperture images, often called sub-aperture images.

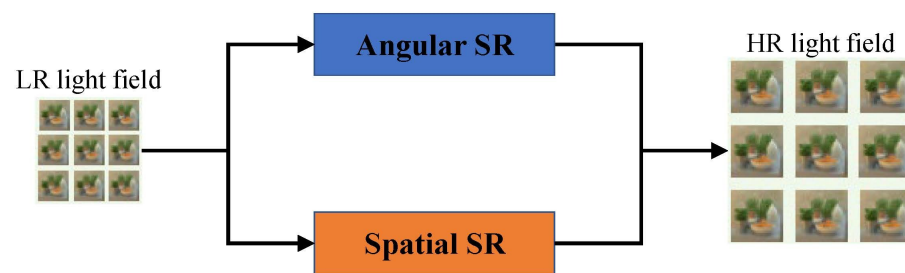
#### 3.1.1. Intra-Image-Similarity-Based LFSR

Early light field super-resolution methods based on deep learning usually divide different sub-tasks for processing, and the results of the sub-tasks work together to generate the final high-resolution light field image.

As shown in Figure 7, the network model proposed in this period usually contains two network branches to process the angular domain and the spatial domain of the light field. The networks designed by Gul [52], Ko [60], and Jin [61] all follow this processing idea. Gul [52] used light field images with low angular resolution and low spatial resolution as the input of the network.

First, through the angular SR network, a new sub-aperture image is synthesized by interpolation and the output has low spatial resolution and high angular resolution. The spatial SR network takes the output of the angular SR network as input, improves the spatial resolution of each sub-aperture image through training, and finally outputs a light field image with high spatial resolution and high angular resolution. Ko [60] designed a hybrid module called AFR (adaptive feature remixing) and embedded it in the spatial SR and angular SR networks.

The AFR module can perform feature remixing on the multi-view features extracted by the network through the disparity estimator according to the angular coordinates. The network can generate high-quality super-resolution images regardless of the angular coordinates of the input view images. The method proposed by Jin [61] used two sub-network modules to model the complementary information between views and the parallax structure of the light field image, while fusing the complementary information between views, the original parallax structure of the light field is preserved.



**Figure 7.** Network model of light field super resolution with two sub-network branches.

In addition to processing the angular domain and the spatial domain of the light field separately, there are also some methods that treat the two as an interconnected whole. Yeung [62] used four-dimensional convolution to characterize the high-dimensional structure of the light field image and designed a special feature extraction layer that can extract the joint spatial and angular features on the light field image to perform super-resolution processing of the light field image.

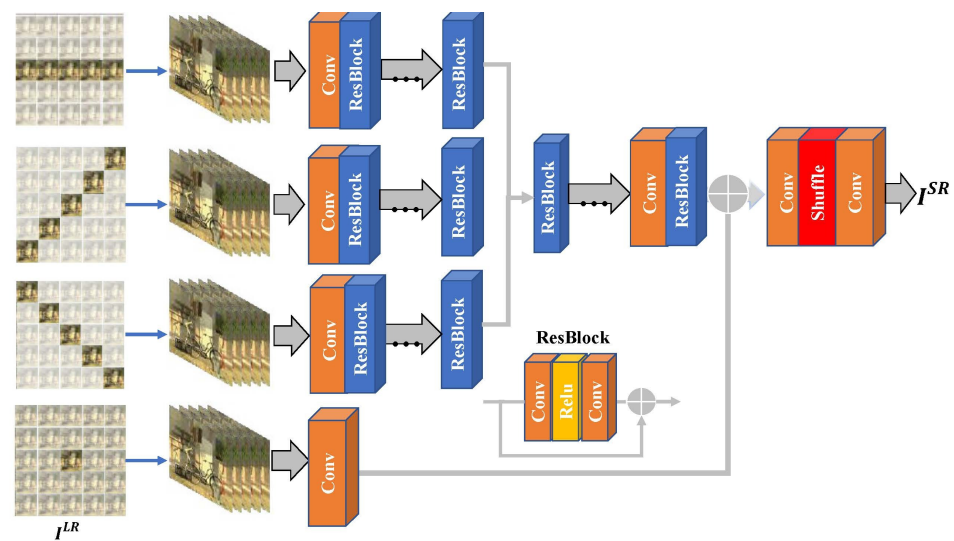
Wang [57] proposed a spatial-angular interactive network, which uses two special convolutional layers to extract space and angle features from the input light field image, and then repeats this process to gradually merge the space and angle information. Finally, the interactive features are merged to super-resolution each sub-aperture image.

Li [63] believes that the spatial and angular information provided by other views in the light field image are not of the same importance, therefore they proposed a light field spatial-angular attention module (LFSAA) to adjust the weight of spatial and angular information in the spatial and angular domains of the light field image, to improve the resolution of the spatial and angular domains with trade-offs.

There are also some methods that only deal with angular resolution or spatial resolution of the light field images, These methods strive to achieve the best local performance. The network proposed by Wang [64] can simultaneously up-sample all sub-aperture images to directly output light field images with high angular resolution. Jin [65] designed a hybrid module to use the parallax geometry information of the light field and reconstruct the light field with high angular resolution based on this information.

Zhang [66] used multiple network branches to learn the sub-pixel details of the corresponding spatial direction from different angles and integrated them for spatial up-sampling to obtain light field images with high spatial resolution. As shown in Figure 8, Zhang [24] explored inherent correspondence of parallax information reflected in the view and learned sub-pixel mapping for different directions. After combining residual information from different spatial directions, super-resolution images with full details were generated.

Influenced by [24], Kim [67] adopted modified residual block to avoid gradient vanishing or exploding and proposed an end-to-end residual network to improve the angular resolution of the light field. Based on view synthesis technology, Kalantari [68] proposed a learning-based method to synthesize new high-resolution views from a set of low-resolution input views. This method sacrifices angular resolution to improve spatial resolution. Ribeiro [69] used deformable convolution to extract angular features, and the extracted angular features were used for feature alignment, which reduces the complexity while improving the quality of the reconstructed image.



**Figure 8.** Network structure proposed by Zhang [24].

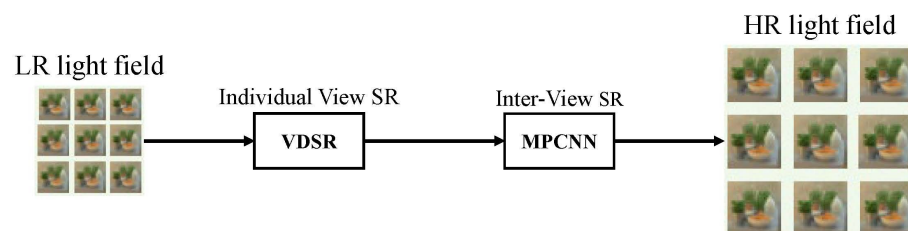
### 3.1.2. Inter-Image-Similarity-Based LFSR

Ordinary image SR based on deep learning tends to exploit only the external phase between images, i.e., training the network with many image datasets, thus embedding the natural image prior into the neural network model. Although for general image SR, better super-resolution performance can be obtained by only using the external similarity of the image; however, this is not sufficient for processing complex light field SR. There is also a high degree of similarity between different angle views in the light field, i.e., the internal similarity of the light field.

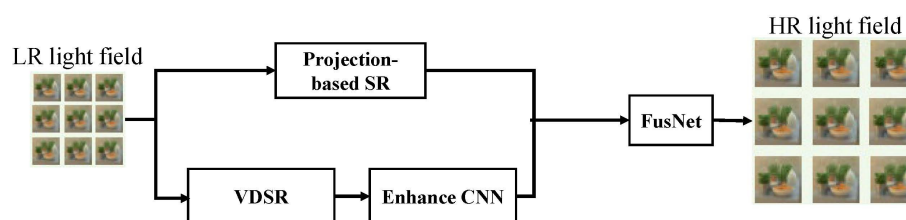
The internal similarity of the light field provides a wealth of information for super-resolution of each view. Therefore, comprehensive utilization of the internal and external similarities of the light field can greatly improve the performance of the learning-based light field SR.

As shown in Figure 9, Fan [53] divides the light field super-resolution processing into a two-stage task, using external similarity and internal similarity in the two stages of the task. In the first stage, the VDSR network is trained to use the external similarity to enhance the view, and in the second stage, the max-pooling convolutional neural network (MPCNN) is trained so that it can use the internal similarity to further enhance the target view from the information of the neighboring views.

As shown in Figure 10, Cheng [70] utilized the internal similarity of the image by introducing the intensity consistency check standard and the back-projection refinement, while the external correlation is learned by the CNN-based method. This method takes the VDSR result of each sub-aperture image as input, and integrates the internal and external similarity of the image.



**Figure 9.** Network structure proposed by Fan [53].



**Figure 10.** Network structure proposed by Cheng [70].

Ma [58] used the most advanced technology of SISR to handle the light field super-resolution task and developed a flexible light field super-resolution network model based on the RDN network. Cao [71] used several different models to deal with the light field image angle super-resolution task under different image forms, and then merged these models to make full use of the light field image's own information. Guo [72] indicated that light rays distributed in different light fields have the same constraint under certain conditions and used the residual network to reduce the error-prone constraints.

Jin [73] propagated spatial details in a single high resolution (HR) view to other views while maintaining the intrinsic structure between the sub-aperture images of the original low resolution (LR) view, and then placed HR view together with the LR view into a network using the spatial-angular interleaved convolutional layer of their design for processing. Cheng [59] proposed a zero-sample learning framework for light field SR, using features extracted from the input low-resolution light field itself to super-resolution the target view. During training, Cheng divided the LFSR task into several sub-tasks and completes them separately.

Jin [74] learned information related to the scene geometry from the sparsely sampled light field and synthesized sub-aperture images from sparsely sampled light field into a new sub-aperture image. Sub-aperture images in sparsely sampled LF and newly synthesized Sub-aperture images together synthesized densely-sampled LF. Newly generated densely-sampled LF were used as the starting point for up-sampling and were reconstructed by residual learning to generate a densely sampled light field. This approach significantly improves the performance of the super-resolution as well as its operational efficiency.

Although the combined use of the internal and external similarities of the light field can bring better super-resolution quality, the network model constructed in this way is more complicated, difficult to train, and inconvenient to use. Therefore, some researchers believe that only by analyzing the internal connections between the sub-aperture images of the light field, the super-resolution quality can be improved while considering the efficiency.

For this consideration, Wang [23] proposed a bidirectional cyclic convolutional neural network embedded with an implicit multi-scale fusion layer to iteratively model the spatial correlation between adjacent sub-aperture images of LF data. Farrugia [75] embeds low-rank priors into the deep convolutional network to restore the consistency of the entire light field on all sub-aperture images. Meng [56] merged the high-dimensional convolutional layer for the special structure of light field images into GAN [76] to find the correlation between adjacent light field images.

### 3.2. Epipolar-Plane-Image-Based LFSR

EPI is a 2D slice of a 4D light field with a constant angle and spatial direction, which contains the depth information of the scene; therefore, it is usually used for the depth estimation of the light field; however, some researchers attempt to apply it to light field super-resolution tasks.

Wanner [48] analyzed the EPI to estimate the disparity map locally, and the obtained disparity map was further used in the super-resolution of the light field image. Wu [27] used the clear texture structure of the EPI in the light field data to model the light field reconstruction problem as a CNN-based EPI angle detail recovery problem. Zhu [55] designed a CNN-LSTM network to capture the continuity of EPI, which can simultaneously super-resolution the spatial and angular dimensions of the image. Meng [77] added high-

order residual blocks to the network so that the network model can extract representative geometric features through global residual learning. These features were used for the spatial angle up-sampling of the EPI.

Liu [78] used the spatial correlation within the LF view and the angle correlation between adjacent LF views to jointly reflect the LF image structure, and proposed a stack representation of the polar plane image volume for LF angle super-resolution. The EPI-VS data was used as the input of the LF angle super-resolution network. Zhao [79] proposed a multi-scale dense residual network to achieve EPI super-resolution and quality enhancement.

Wafa [80] designed an end-to-end deep-learning model to process all sub-aperture images at the same time, and used EPI information to smoothly generate views. Yuan [54] used EPI to restore the geometric consistency of light field images lost in SISR processing and proposed a network framework consisting of SISR deep CNN and EPI enhanced deep CNN. Inspired by the non-local attention mechanism [81], Wu [82] computed attention non-locally on the epipolar plane pixel by pixel, thus generating an attention map of the spatial dimension and guiding the reconstruction of the corresponding angular dimension based on the generated attention map.

#### 4. Data Set and Comparison

##### 4.1. Data Set

In chronological order, the current main light field data sets available for training and testing include: HCI old [49], STFlytro [83], EPFL [51], HCI [50], 30scenes [68]. Among them, HCI old, HCI, and 30scenes belong to the synthetic image data set, and the images of STFlytro and EPFL come from real-world images collected by a camera. The data set list is shown in Table 2.

**Table 2.** Overview of the light field super-resolution data sets.

Data Set	Years	Number of Scenes	Shooting Method
HCI old [49]	2013	13	Blender Synthesis
STFlytro [83]	2016	9	Lytro Illum
EPFL [51]	2016	10	Lytro Illum
HCI [50]	2016	24	Blender Synthesis
30scenes [68]	2016	30	CNN Synthesis

##### 4.2. Comparison

Table 3 shows the comparison between traditional method and deep-learning-based method. Traditional methods are mainly based on expert experience and prior knowledge, which can achieve better reconstruction quality at local details; however, the overall quality is sacrificed. Deep-learning-based methods can automatically reconstruct image by training network over huge amount of data, and the reconstructed image usually has a quality improvement at both local and global scale. In addition, compared with traditional methods, deep-learning-based method has faster processing speed when faced with a large batch of LSFR tasks.

**Table 3.** Comparison of traditional methods and deep-learning-based methods for light field image super resolution.

	Traditional Method	Deep-Learning-Based Method
Reconstruction Quality	Good detail but poor overall quality	Good detail and overall quality
Advantages	No training required. Process explainable.	Automatic feature extraction. Parallel processing.
Disadvantages	Relying on expert experience. Weak generalization ability. Poor robustness.	High computational complexity. Relying on dataset.

As for performance, several traditional and deep-learning-based LFSR works are selected for comparison, as shown in Table 4. The  $\times 2$  SR ratio is chosen. PSNR and SSIM are evaluation metrics.

1. The performance of traditional methods, such as [34,35,40,48] on the HCI old data set is incomparable with the methods based on deep learning [22]. A recent work [42] incorporates multiple optical factors that may affect image quality into error considerations, and performs image reconstruction and demosaicing jointly in an optimized framework, resulting in a significant improvement in the quality of the reconstructed image. Nevertheless, traditional-based methods still cannot compete with deep-learning-based methods in terms of performance.
2. For synthetic image data sets, such as HCI old and HCI, methods based on deep learning, such as [23,24,60,70] have good performance; however, for data sets taken from the real world, such as EPFL, the processing performance is significantly reduced. The reason for this result is that real-world light field images often have a lot of noise that affects imaging when shooting, such as picture shake, light interference, etc.; thus, it is more difficult to process real light field images compared with synthesized light field images.
3. Compared with the network model designed for SISR, such as [45], the network designed for the special high-dimensional structure of the light field effectively improves the performance, such as [61,62]. The network model that comprehensively considers the spatial resolution and angular resolution of the light field also has good results, especially [57] maintains a high PSNR level on multiple data sets.
4. Refs. [53,58] demonstrate that advanced SISR techniques are highly informative for LFSR work. Although refs. [53,58] achieved high performance on the test set; however, ref. [53] used two super-resolution networks for the light field images, which increases the processing time significantly. while ref. [58] allows the network designed for SISR to process the additional angular information in the light field images by varying the size of the convolution kernel, which also increases the computational complexity.
5. EPI-based LFSR methods have superior performance on datasets consisting of real-world images, such as [27,80]. This is because the depth information of the scene is contained in the EPI map, and using this depth information can significantly improve the super-resolution performance for real-world light field images.
6. Residual learning has been proven to greatly improve the performance of SISR, and the introduction of residual learning into LFSR tasks also has good performance, such as [67,74,77]. However, the dataset selected during training seems to have some influence on the final super-resolution results. Refs. [67,77] both selected real-world images taken by Lytro Illum during training, ref. [74] chose to use both synthetic LF images and real LF images for training. In tests, ref. [77] had good performance on real-world image datasets, such as EPFL, while for synthetic datasets: HCI old, HCI, its performance significantly decreased, ref. [74] maintained high performance on both synthetic and real-world datasets.
7. As a widely noticed network model, GAN has been used in single-image super-resolution, and in super-resolution tasks for light field images, the LightGAN proposed by [27] has a good performance on the EPFL, higher than [23,70], although it has a lower performance on the HCI. There is still much room for improvement using GAN for light field super-resolution tasks, and it is worthwhile to investigate how to better combine GAN with light fields.
8. Transformer [84], which has made a big splash in NLP in recent years, has also provided new inspiration for super-resolution of light field images. The successful attempt of [63,82] demonstrates the feasibility of using the attention mechanism for super-resolution tasks in light fields, where the introduction of the attention mechanism can better extract the additional angular information contained in the light field image, thus improving the super-resolution quality.

**Table 4.** Performance comparison of light field super-resolution algorithms, where best results are in bold and “-” means not tested. The five methods above are traditional methods, while others are deep learning based methods.

Method	Dataset	HCI Old (PSNR/SSIM)	HCI (PSNR/SSIM)	EPFL (PSNR/SSIM)	STF Lytro (PSNR/SSIM)
Mitra [35]		29.60/0.899	-	-	25.70/0.724
Wanner [48]		30.22/0.901	-	-	-
Wang [34]		35.14/0.951	-	-	-
Farrugia [40]		30.57/-	-	-	32.13/-
Pendu [42]		38.64/-	36.77/-	-	-
Yoon [22]		37.47/0.974	-	-	29.50/0.796
Wang [23]		36.46/0.964	33.63/0.932	32.70/0.935	30.31/0.815
Zhang [24]		41.09/0.988	36.45/0.979	35.48/0.973	-
Kim [45]		40.34/0.985	34.37/0.956	32.01/0.959	29.99/0.803
Ko [60]		42.06/0.989	37.21/0.977	36.00/0.982 -	-
Jin [61]		-	38.52/0.959	-	41.96/0.979
Yeung [62]		-	-	-	40.50/0.977
Wang [57]		44.65/0.995	37.20/0.976	34.76/0.976	38.81/0.983
Zhang [66]		42.14/0.981	37.01/0.963	35.81/0.961	-
Fan [53]		40.77/0.968	-	-	-
Cheng [70]		36.10/-	-	30.41/-	-
Ma [58]		43.90/0.993	<b>40.49/0.986</b>	<b>41.38/0.989</b>	-
Jin [73]		-	-	-	34.39/0.951
Cheng [59]		40.03/-	37.94/-	34.78/-	38.05/-
Ribeiro [69]		<b>45.49/0.964</b>	38.22/0.956	34.41/0.953	-
Farrugia [75]		-	-	-	32.41/0.884
Meng [56]		-	32.45/-	34.20/-	-
Wu [27]		-	-	-	42.48/-
Zhu [55]		-	-	-	33.04/0.958
Wafa [80]		39.76/0.968	-	-	<b>44.45/0.995</b>
Yuan [54]		38.63/0.954	-	-	40.61/0.984
Meng [77]		33.12/0.913	34.64/0.933	35.97/0.947	38.30/0.969
Kim [67]		-	-	-	39.25/0.990
Jin [74]		41.80/0.974	37.14/0.966	-	-

## 5. Existing Challenges and Future Developments

Although the performance of light field image super-resolution is constantly improving, there are still many difficulties that need to be resolved. Listed below are challenges in the current light field super-resolution field.

### 5.1. Existing Challenges

Although the development of deep learning has promoted the research of light field super-resolution processing, the network model trained from a self-defined light field data set is often not universal. Therefore, a unified test environment, i.e., a benchmark data set, is needed for the comparison of different light field super-resolution methods. Compared with 2D planar images or 3D stereoscopic images, 4D light field images have high-dimensional characteristics, and thus the collection of light field images requires special equipment. The camera array is one of the main ways to capture light field data.

Although this collection method decouples light by fixing multiple camera lenses to reduce the resolution loss of imaging; however, the shooting system is complicated, the equipment cost is high and the land area is large. The difficulty of light field acquisition has led to the lack of benchmark data sets that can be used for experiments. The spatial resolution and the angular resolution of the light field are a pair of interrelated but contradictory factors. Compared with 2D planar images, the light field image contains additional angle information, which can be used to calculate images with different depths of field; however, this is at the expense of the spatial resolution of the picture.

There is a microprism matrix in the light field camera. Each microprism in the matrix can be regarded as a small camera. When shooting, it is equivalent to having multiple cameras shooting from different angles at the same time, to collect light from different perspectives. Light field images are imaged at the same resolution; however, when imaging, the light field camera sacrifices the spatial resolution of “each small camera” to obtain angular information, which is one of the reasons why the image obtained from light field imaging is less pixelated and much blurry than ordinary imaging. How to improve the spatial resolution of the light field while ensuring the angular resolution of the light field is also a major research challenge.

In addition to the above mentioned, the data volume of light field is much larger than traditional image/video data, causing problems in storage, transmission and compression. In terms of light field super-resolution, the impact of data size on super-resolution is obvious. For deep-learning-based light field super-resolution tasks, too large data volume can seriously slow down the efficiency of training and make it more difficult for each layer of the network to process the data.

### 5.2. Future Developments

First, the use of deep learning to solve light field super-resolution tasks is effective; however, the particular high-dimensional structure of light field data also poses many difficulties for super-resolution processing. Network models, such as VDSR designed for SISR tasks have been confirmed not to be suitable for super-resolution tasks in light field. Compared with [60,63], these network models do not take into account the special 4D structure of the light field.

The light field images are only processed on a 2D level. This treatment not only loses much of the information between the views of the light field but also does not take full advantage of the angular and spatial characteristics of the light field image. If the high-dimensional structure of the light field is not taken into account, the result of super-resolution processing will not be able to balance the angular resolution with the spatial resolution. The result is that the processed image loses scene angle information or the spatial resolution is not significantly improved. Therefore, the development of network models for light field structures is of great importance to improve the performance of light field image super-resolution.

Current single-image super-resolution techniques based on deep learning can compress the processing time of low-complexity images to around 1 second, while the processing time of high-complexity images varies from a few seconds to a dozen seconds. Compared to ordinary 2D flat images, light field images with high-dimensional data structures can be considered to be generally high complexity data, requiring deeper networks to be built to process the light field images, which takes more time. Therefore, ensuring super-resolution quality while keeping the network lightweight is also a worthwhile consideration.

In addition, the lack of evaluation metrics for light field image quality assessment is also an issue worth considering. The current evaluation metrics for light field image quality are usually PSNR and SSIM, which do not take into account the additional depth of field information contained in the light field image. It is also worthwhile to evaluate the reconstruction of the depth of field in the super-resolution processed light field image. If an evaluation index can be designed for the light field depth information, it is possible to make better use of the depth information contained in the light field image for super-resolution processing.

## 6. Conclusions

This paper introduced the concept of light field and light field super-resolution, listed traditional light field super-resolution techniques and deep-learning-based light field super-resolution techniques, and mainly compares and summarizes the deep-learning-based light field super-resolution techniques.

The comparison shows that advanced network models and more complete datasets can significantly improve the performance of light field super-resolution, while convolutional layers designed for the high-dimensional data characteristics of light fields can also bring significant performance improvements.

Future research on light field super-resolution techniques should focus on the design of the network structure, it is worth considering how to adapt the network structure to the high-dimensional nature of the light field data.

**Author Contributions:** Conceptualization, L.Y. and K.C.; methodology, L.Y.; validation, S.H., K.C. and L.Y.; investigation, Y.M.; writing—original draft preparation, L.Y. and Y.M.; writing—review and editing, L.Y.; visualization, Y.M.; funding acquisition, L.Y. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported in part by the National Natural Science Foundation of China under Grant 62002172; and in part by the Natural Science Foundation of the Jiangsu Higher Education Institutions of China under Grant 19KJB510040; and in part by the Nanjing Scientific Innovation Foundation for the Returned Overseas Chinese Scholars under Grant R2019LZ04; and in part by the Jiangsu Provincial Double-Innovation Doctor Program; and in part by the Open Project Program of the National Laboratory of Pattern Recognition (NLPR) under Grant 202100002; and in part by the Startup Foundation for Introducing Talent of NUIST under Grant 2018r080.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

- Faraday, M. LIV. Thoughts on ray-vibrations. *Lond. Edinb. Dublin Philos. Mag. J. Sci.* **1846**, *28*, 345–350. [\[CrossRef\]](#)
- Gershun, A. The light field. *J. Math. Phys.* **1939**, *18*, 51–151. [\[CrossRef\]](#)
- Bergen, J.R.; Adelson, E.H. The plenoptic function and the elements of early vision. *Comput. Model. Vis. Process.* **1991**, *1*, 8.
- Levoy, M.; Hanrahan, P. Light field rendering. In Proceedings of the 23rd annual conference on Computer graphics and interactive techniques, New York, NY, USA, 4–9 August 1996; pp. 31–42.
- Georgiev, T.; Yu, Z.; Lumsdaine, A.; Goma, S. Lytro camera technology: Theory, algorithms, performance analysis. *Multimed. Content Mob. Devices* **2013**, *8667*, 458–467.
- Guillo, L.; Jiang, X.; Lafruit, G.; Guillemot, C. *Light Field Video Dataset Captured by a R8 Raytrix Camera (with Disparity Maps)*; International Organisation for Standardisation ISO/IEC JTC1/SC29/WG1 & WG11; ISO: Geneva, Switzerland, 2018.
- Yang, J.C.; Everett, M.; Buehler, C.; McMillan, L. A real-time distributed light field camera. *Render. Tech.* **2002**, *2002*, 77–86.
- Venkataraman, K.; Lelescu, D.; Duparré, J.; McMahon, A.; Molina, G.; Chatterjee, P.; Mullis, R.; Nayar, S. Picam: An ultra-thin high performance monolithic camera array. *ACM Trans. Graph. (TOG)* **2013**, *32*, 1–13. [\[CrossRef\]](#)
- Levin, A.; Fergus, R.; Durand, F.; Freeman, W.T. Image and depth from a conventional camera with a coded aperture. *ACM Trans. Graph. (TOG)* **2007**, *26*, 70-es. [\[CrossRef\]](#)
- Inagaki, Y.; Kobayashi, Y.; Takahashi, K.; Fujii, T.; Nagahara, H. Learning to capture light fields through a coded aperture camera. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 418–434.
- 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.
- Irani, M.; Peleg, S. Improving resolution by image registration. *CVGIP Graph. Model. Image Process.* **1991**, *53*, 231–239. [\[CrossRef\]](#)
- Nava, F.P.; Luke, J. Simultaneous estimation of super-resolved depth and all-in-focus images from a plenoptic camera. In Proceedings of the 2009 3DTV Conference: The True Vision-Capture, Transmission and Display of 3D Video, Potsdam, Germany, 4–6 May 2009; pp. 1–4.
- 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. [\[CrossRef\]](#)
- Wang, Z.; Chen, J.; Hoi, S.C. Deep learning for image super-resolution: A survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *43*, 3365–3387. [\[CrossRef\]](#)
- Eigen, D.; Puhrsch, C.; Fergus, R. Depth map prediction from a single image using a multi-scale deep network. In Proceedings of the Advances in Neural Information Processing Systems, Montreal, QC, Canada, 8–13 December 2014; Volume 27.
- Wang, X.; Chen, S.; Liu, J.; Wei, G. High Edge-Quality Light-Field Salient Object Detection Using Convolutional Neural Network. *Electronics* **2022**, *11*, 1054. [\[CrossRef\]](#)

18. Adjabi, I.; Ouahabi, A.; Benzaoui, A.; Taleb-Ahmed, A. Past, present, and future of face recognition: A review. *Electronics* **2020**, *9*, 1188. [\[CrossRef\]](#)
19. Wang, M.; Deng, W. Deep face recognition: A survey. *Neurocomputing* **2021**, *429*, 215–244. [\[CrossRef\]](#)
20. Minaee, S.; Abdolrashidi, A.; Su, H.; Bennamoun, M.; Zhang, D. Biometrics recognition using deep learning: A survey. *arXiv* **2019**, arXiv:1912.00271.
21. Khaldi, Y.; Benzaoui, A.; Ouahabi, A.; Jacques, S.; Taleb-Ahmed, A. Ear Recognition Based on Deep Unsupervised Active Learning. *IEEE Sens. J.* **2021**, *21*, 20704–20713. doi: 10.1109/JSEN.2021.3100151. [\[CrossRef\]](#)
22. 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, Washington, DC, USA, 7–13 December 2015; pp. 24–32.
23. Wang, Y.; Liu, F.; Zhang, K.; Hou, G.; Sun, Z.; Tan, T. LFNet: A novel bidirectional recurrent convolutional neural network for light-field image super-resolution. *IEEE Trans. Image Process.* **2018**, *27*, 4274–4286. [\[CrossRef\]](#) [\[PubMed\]](#)
24. 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.
25. Guillemot, C.; Farrugia, R.A. Light field image processing: Overview and research issues. *IEEE J. Sel. Top. Signal Process.* **2017**, *11*, 926–954.
26. Ihrke, I.; Restrepo, J.; Mignard-Debise, L. Principles of light field imaging: Briefly revisiting 25 years of research. *IEEE Signal Process. Mag.* **2016**, *33*, 59–69. [\[CrossRef\]](#)
27. Wu, G.; Zhao, M.; Wang, L.; Dai, Q.; Chai, T.; Liu, Y. Light field reconstruction using deep convolutional network on EPI. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 6319–6327.
28. Lim, J.; Ok, H.; Park, B.; Kang, J.; Lee, S. Improving the spatio-temporal resolution based on 4D light field data. In Proceedings of the 2009 16th IEEE International Conference on Image Processing (ICIP), Cairo, Egypt, 7–10 November 2009; pp. 1173–1176.
29. Stark, H.; Oskoui, P. High-resolution image recovery from image-plane arrays, using convex projections. *JOSA A* **1989**, *6*, 1715–1726. [\[CrossRef\]](#)
30. Ng, R. Fourier slice photography. In *ACM Siggraph 2005 Papers*; ACM: New York, NY, USA, 2005; pp. 735–744.
31. Pérez, F.; Pérez, A.; Rodríguez, M.; Magdaleno, E. Fourier slice super-resolution in plenoptic cameras. In Proceedings of the 2012 IEEE International Conference on Computational Photography (ICCP), Seattle, WA, USA, 28–29 April 2012; pp. 1–11.
32. Yu, Z.; Yu, J.; Lumsdaine, A.; Georgiev, T. An analysis of color demosaicing in plenoptic cameras. In Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition, Washington, DC, USA, 16–21 June 2012; pp. 901–908.
33. Zhou, S.; Yuan, Y.; Su, L.; Ding, X.; Wang, J. Multiframe super resolution reconstruction method based on light field angular images. *Opt. Commun.* **2017**, *404*, 189–195. [\[CrossRef\]](#)
34. Wang, Y.; Hou, G.; Sun, Z.; Wang, Z.; Tan, T. A simple and robust super resolution method for light field images. In Proceedings of the 2016 IEEE International Conference on Image Processing (ICIP), Phoenix, AZ, USA, 25–28 September 2016; pp. 1459–1463.
35. 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, Providence, RI, USA, 16–21 June 2012; pp. 22–28.
36. Rossi, M.; Frossard, P. Graph-based light field super-resolution. In Proceedings of the 2017 IEEE 19th International Workshop on Multimedia Signal Processing (MMSP), Luton, UK, 16–18 October 2017; pp. 1–6.
37. 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.
38. Egiazarian, K.; Katkovnik, V. Single image super-resolution via BM3D sparse coding. In Proceedings of the 2015 23rd European Signal Processing Conference (EUSIPCO), Nice, France, 31 August 31–4 September 2015; pp. 2849–2853.
39. Alain, M.; Smolic, A. Light field denoising by sparse 5D transform domain collaborative filtering. In Proceedings of the 2017 IEEE 19th International Workshop on Multimedia Signal Processing (MMSP), Luton, UK, 16–18 October 2017; pp. 1–6.
40. 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. [\[CrossRef\]](#)
41. Boominathan, V.; Mitra, K.; Veeraraghavan, A. Improving resolution and depth-of-field of light field cameras using a hybrid imaging system. In Proceedings of the 2014 IEEE International Conference on Computational Photography (ICCP), Santa Clara, CA, USA, 2–4 May 2014; pp. 1–10.
42. Le Pendu, M.; Smolic, A. High resolution light field recovery with fourier disparity layer completion, demosaicing, and super-resolution. In Proceedings of the 2020 IEEE International Conference on Computational Photography (ICCP), Saint Louis, MO, USA, 24–26 April 2020; pp. 1–12.
43. Le Pendu, M.; Guillemot, C.; Smolic, A. A fourier disparity layer representation for light fields. *IEEE Trans. Image Process.* **2019**, *28*, 5740–5753. [\[CrossRef\]](#) [\[PubMed\]](#)
44. Dong, C.; Loy, C.C.; He, K.; Tang, X. Image super-resolution using deep convolutional networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *38*, 295–307. [\[CrossRef\]](#) [\[PubMed\]](#)
45. 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, Las Vegas, NV, USA, 27–30 June 2016; pp. 1646–1654.

46. 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.
47. Dong, C.; Loy, C.C.; He, K.; Tang, X. Learning a deep convolutional network for image super-resolution. In Proceedings of the European Conference on Computer Vision, Zurich, Switzerland, 5–12 September 2014; pp. 184–199.
48. Wanner, S.; Goldluecke, B. Variational light field analysis for disparity estimation and super-resolution. *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**, *36*, 606–619. [\[CrossRef\]](#)
49. Wanner, S.; Meister, S.; Goldluecke, B. Datasets and benchmarks for densely sampled 4D light fields. In Proceedings of the Vision, Modeling, and Visualization, Lugano, Switzerland, 11–13 September 2013; Volume 13, pp. 225–226.
50. Honauer, K.; Johannsen, O.; Kondermann, D.; Goldluecke, B. A dataset and evaluation methodology for depth estimation on 4D light fields. In Proceedings of the Asian Conference on Computer Vision, Taipei, Taiwan, China, 20–24 November 2016; pp. 19–34.
51. Rerabek, M.; Ebrahimi, T. New light field image dataset. In Proceedings of the Eighth International Conference on Quality of Multimedia Experience (QoMEX), Lisbon, Portugal, 6–8 June 2016.
52. Gul, M.S.K.; Gunturk, B.K. Spatial and angular resolution enhancement of light fields using convolutional neural networks. *IEEE Trans. Image Process.* **2018**, *27*, 2146–2159. [\[CrossRef\]](#)
53. Fan, H.; Liu, D.; Xiong, Z.; Wu, F. Two-stage convolutional neural network for light field super-resolution. In Proceedings of the 2017 IEEE International Conference on Image Processing (ICIP), Beijing, China, 17–20 September 2017; pp. 1167–1171.
54. Yuan, Y.; Cao, Z.; Su, L. Light-field image superresolution using a combined deep CNN based on EPI. *IEEE Signal Process. Lett.* **2018**, *25*, 1359–1363. [\[CrossRef\]](#)
55. Zhu, H.; Guo, M.; Li, H.; Wang, Q.; Robles-Kelly, A. Breaking the spatio-angular trade-off for light field super-resolution via lstm modelling on epipolar plane images. *arXiv* **2019**, arXiv:1902.05672.
56. Meng, N.; Ge, Z.; Zeng, T.; Lam, E.Y. LightGAN: A deep generative model for light field reconstruction. *IEEE Access* **2020**, *8*, 116052–116063. [\[CrossRef\]](#)
57. 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 European Conference on Computer Vision, Glasgow, UK, 23–28 August 2020; pp. 290–308.
58. Ma, D.; Lumsdaine, A.; Zhou, W. Flexible Spatial and Angular Light Field Super Resolution. In Proceedings of the 2020 IEEE International Conference on Image Processing (ICIP), Anchorage, AK, USA, 25–28 October 2020; pp. 2970–2974.
59. Cheng, Z.; Xiong, Z.; Chen, C.; Liu, D.; Zha, Z.J. Light field super-resolution with zero-shot learning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Virtual, 20–25 June 2021; pp. 10010–10019.
60. Ko, K.; Koh, Y.J.; Chang, S.; Kim, C.S. Light field super-resolution via adaptive feature remixing. *IEEE Trans. Image Process.* **2021**, *30*, 4114–4128. [\[CrossRef\]](#)
61. 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, Virtual, 13–19 June 2020; pp. 2260–2269.
62. Yeung, H.W.F.; Hou, J.; Chen, X.; Chen, J.; Chen, Z.; Chung, Y.Y. Light field spatial super-resolution using deep efficient spatial-angular separable convolution. *IEEE Trans. Image Process.* **2018**, *28*, 2319–2330. [\[CrossRef\]](#)
63. Li, D.; Yang, D.; Wang, S.; Sheng, H. Light Field Super-Resolution Based on Spatial and Angular Attention. In Proceedings of the International Conference on Wireless Algorithms, Systems, and Applications, Nanjing, China, 25–27 June 2021; pp. 314–325.
64. Wang, X.; You, S.; Zan, Y.; Deng, Y. Fast light field angular resolution enhancement using convolutional neural network. *IEEE Access* **2021**, *9*, 30216–30224. [\[CrossRef\]](#)
65. Jin, J.; Hou, J.; Yuan, H.; Kwong, S. Learning light field angular super-resolution via a geometry-aware network. In Proceedings of the AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020; Volume 34, pp. 11141–11148.
66. 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. [\[CrossRef\]](#) [\[PubMed\]](#)
67. Kim, D.M.; Kang, H.S.; Hong, J.E.; Suh, J.W. Light field angular super-resolution using convolutional neural network with residual network. In Proceedings of the 2019 Eleventh International Conference on Ubiquitous and Future Networks (ICUFN), Split, Croatia, 2–5 July 2019; pp. 595–597.
68. Kalantari, N.K.; Wang, T.C.; Ramamoorthi, R. Learning-based view synthesis for light field cameras. *ACM Trans. Graph. (TOG)* **2016**, *35*, 1–10. [\[CrossRef\]](#)
69. Ribeiro, D.A.; Silva, J.C.; Lopes Rosa, R.; Saadi, M.; Mumtaz, S.; Wuttisittikulkij, L.; Zegarra Rodriguez, D.; Al Otaibi, S. Light field image quality enhancement by a lightweight deformable deep learning framework for intelligent transportation systems. *Electronics* **2021**, *10*, 1136. [\[CrossRef\]](#)
70. Cheng, Z.; Xiong, Z.; Liu, D. Light field super-resolution by jointly exploiting internal and external similarities. *IEEE Trans. Circuits Syst. Video Technol.* **2019**, *30*, 2604–2616. [\[CrossRef\]](#)
71. Cao, F.; An, P.; Huang, X.; Yang, C.; Wu, Q. Multi-Models Fusion for Light Field Angular Super-Resolution. 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. 2365–2369.
72. Guo, M.; Zhu, H.; Zhou, G.; Wang, Q. Dense light field reconstruction from sparse sampling using residual network. In Proceedings of the Asian Conference on Computer Vision, Perth, Australia, 2–6 December 2018; pp. 50–65.

73. Jin, J.; Hou, J.; Chen, J.; Yeung, H.; Kwong, S. Light Field Spatial Super-resolution via CNN Guided by A Single High-resolution RGB Image. In Proceedings of the 2018 IEEE 23rd International Conference on Digital Signal Processing (DSP), Shanghai, China, 19–21 November 2018; pp. 1–5.
74. Jin, J.; Hou, J.; Chen, J.; Zeng, H.; Kwong, S.; Yu, J. Deep coarse-to-fine dense light field reconstruction with flexible sampling and geometry-aware fusion. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *44*, 1819–1836. [[CrossRef](#)]
75. Farrugia, R.A.; Guillemot, C. Light field super-resolution using a low-rank prior and deep convolutional neural networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2019**, *42*, 1162–1175. [[CrossRef](#)]
76. 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.
77. Meng, N.; Thus, H.K.H.; Sun, X.; Lam, E.Y. High-dimensional dense residual convolutional neural network for light field reconstruction. *IEEE Trans. Pattern Anal. Mach. Intell.* **2019**, *43*, 873–886. [[CrossRef](#)]
78. Liu, D.; Wu, Q.; Huang, Y.; Huang, X.; An, P. Learning from EPI-Volume-Stack for Light Field image angular super-resolution. *Signal Process. Image Commun.* **2021**, *97*, 116353. [[CrossRef](#)]
79. Zhao, J.; An, P.; Huang, X.; Yang, C.; Shen, L. Light field image compression via CNN-based EPI super-resolution and decoder-side quality enhancement. *IEEE Access* **2019**, *7*, 135982–135998. [[CrossRef](#)]
80. Wafa, A.; Pourazad, M.T.; Nasiopoulos, P. A deep learning based spatial super-resolution approach for light field content. *IEEE Access* **2020**, *9*, 2080–2092. [[CrossRef](#)]
81. Wang, X.; Girshick, R.; Gupta, A.; He, K. Non-local neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 7794–7803.
82. Wu, G.; Wang, Y.; Liu, Y.; Fang, L.; Chai, T. Spatial-angular attention network for light field reconstruction. *IEEE Trans. Image Process.* **2021**, *30*, 8999–9013. [[CrossRef](#)] [[PubMed](#)]
83. Raj, A.S.; Lowney, M.; Shah, R. *Light-Field Database Creation and Depth Estimation*; Stanford University: Palo Alto, CA, USA, 2016.
84. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 5998–6008.