


Article

Light Field Spatial Super-Resolution via View Interaction and the Fusing of Hierarchical Features

Wei Wei ¹, Qian Zhang ^{1,*}, Chunli Meng ^{1,*}, Bin Wang ¹, Yun Fang ² and Tao Yan ³

¹ College of Information, Mechanical and Electrical Engineering, Shanghai Normal University, Shanghai 201418, China; 1000467199@smail.shnu.edu.cn (W.W.); binwang@shnu.edu.cn (B.W.)

² Mathematics & Science College, Shanghai Normal University, Shanghai 200234, China; fangyun0919@shnu.edu.cn

³ School of Mechanical, Electrical & Information Engineering, Putian University, Putian 351100, China; yantao@ptu.edu.cn

* Correspondence: qianzhang@shnu.edu.cn (Q.Z.); windymeng@shnu.edu.cn (C.M.)

Abstract: Light field (LF) cameras can capture the intensity and angle information of any scene in one single shot, and are widely used in virtual reality, refocusing, de-occlusion, depth estimation, etc. However, the fundamental limitation between angular and spatial resolution leads to low spatial resolution of LF images, which limits their development prospects in various fields. In this paper, we propose a new super-resolution network based on view interaction and hierarchical feature fusion to effectively improve LF image spatial resolution and preserve the consistency of the reconstructed LF structure. Initially, we provide novel view interaction blocks to represent the relationship between all views, and efficiently combine hierarchical features using a feature fusion block consisting of residual dense blocks (RDBs) to more effectively preserve the parallax structure. In addition, in the process of extracting features, we introduce residual channel-reconstruction blocks (RCBs) to minimize the duplication of channels in the features. An inter-view unit (InterU) and an intra-view unit (IntraU) are employed to minimize redundancy in the spatial domain as well. Our suggested method is tested using public LF datasets, which include large-disparity LF images. The experimental results demonstrate that our method exhibits the best performance in terms of both quantitative and qualitative results.



Citation: Wei, W.; Zhang, Q.; Meng, C.; Wang, B.; Fang, Y.; Yan, T. Light Field Spatial Super-Resolution via View Interaction and the Fusing of Hierarchical Features. *Electronics* **2024**, *13*, 2373. <https://doi.org/10.3390/electronics13122373>

Academic Editor: Carlo Petrarca

Received: 19 May 2024

Revised: 8 June 2024

Accepted: 14 June 2024

Published: 17 June 2024



Copyright: © 2024 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/>).

Keywords: light field image; spatial super-resolution; residual channel-reconstruction block; hierarchical feature fusion

1. Introduction

LF cameras can simultaneously capture angular information and spatial information because of the insertion of a micro-lens array between the imaging sensor and the main lens to reconstruct multi-view images of a scene [1], that is, encoding three-dimensional (3D) scenes into 4D LF images. The wealth of information contained in these images facilitates many useful applications, such as post-capture refocusing [2], de-occlusion [3], saliency detection [4], depth estimation [5] and virtual reality [6]. However, there is a fundamental limitation between angular resolution and spatial resolution in the image sampling process of LF cameras. In other words, in order to obtain high-resolution sub-aperture images (SAIs), only sparse angular sampling can be performed. In order to achieve dense angular sampling, the spatial resolution of LF images must be reduced. As a result, light field super-resolution (LFSR) can be subdivided into three types: spatial super-resolution (SSR), angular super-resolution (ASR) and spatial-angular super-resolution (SASR). The research focus of this paper is on the SSR problem.

Most SSR approaches are subdivided into two categories: optimization-based approaches and learning-based approaches. The aim of the optimization-based approach is to physically model the relationship between views with estimated disparity information,

and formulate the SR problem as an optimization problem. This is a traditional LF image SR method that explicitly records SAIs using the estimated differences. Alain et al. [7] used LFBM5D sparse coding to implement SSR, and they transformed the SSR problem into an optimization problem based on sparse priors. Rossi et al. [8] proposed an SSR network based on graphical regularization to transform the SR problem into a global optimization problem. Ghassab et al. [9] introduced an LFSR method based on edge-preserving map regularization, which reduces the LF reconstruction error by using an enhanced ADMM model. However, most of the existing disparity estimation methods have noise, occlusion and untextured areas [10], which leads to obvious artifacts in the reconstructed LF images.

In recent years, there has been a gradual rise in learning-based methods, which use complementary information between all SAIs to learn to perform mapping from low-resolution (LR) images to high-resolution (HR) images. These are methods that can learn the LF geometry implicitly. Most learning-based approaches solve the 4D LF SSR problem by exploring complementary information between SAIs through data-driven training. For example, Jin et al. [11] presented a novel LFSR framework for SSR through deep-combination geometric embedding and structural consistency regularization. Zhang et al. [12] suggested a learning-based residual convolutional method to accomplish LF SSR, which can achieve LFSR from different angular resolutions. Zhou et al. [13] put forward a disentangled feature distillation network for LF SSR with degradations. Wang et al. [14] proposed a spatial-angular interaction framework for LFSR. It takes full advantage of the information within each image and the information between all images to obtain HR SAIs. Wang et al. [15] performed SR through deformable convolution. Cheng et al. [16] achieved SSR by combining external and internal similarities. Park et al. [17] presented a multilevel-based LF SSR method to efficiently estimate and mix subpixel information in adjacent images. Although those approaches have greatly promoted efficiency and performance, there are still two key issues that remain unresolved. The first one is that the complementary information between all SAIs is not fully used. The second one is that the consistency of the reconstructed LF structure is not well reserved.

Therefore, in order to solve these two key problems, we propose a new super-resolution network based on view interaction and hierarchical feature fusion. The network is mainly composed of three key components: feature extraction blocks, view interaction blocks and a reconstruction block. The contribution of this paper can be summarized in the following three points:

- (1) We propose residual channel-reconstruction blocks (RCBs) to reduce channel redundancy between features and effectively perform feature extraction in feature extraction blocks. In addition, we use residual atrous spatial pyramid pooling (ResASPP) to enlarge the receptive field and capture information at several scales.
- (2) We introduce InterU and IntraU to take full advantage of the complementary information between all SAIs. At the same time, we further reduce redundancy in the spatial domain by introducing a spatial reconstruction unit (SRU) in view interaction blocks. Our multi-view aggregation block (MAB) is a long-term correlation model based on extended 3D convolution, which further improves the performance of SR networks.
- (3) We effectively fuse shallow features and deep features to retain the consistency of the reconstructed LF structure to obtain high reconstruction accuracy in the reconstruction block. We also perform a number of experiments to demonstrate that our network can achieve state-of-the-art performance beyond several existing excellent methods. Through a complete ablation study, we discuss the importance of the components proposed in this article to give an insight into effective SSR.

The rest of this paper is organized as follows: Firstly, Section 2 gives a brief introduction about LF representations and three different types of LF image super-resolution studies. Section 3 elaborates on each sub-module of the proposed LF SSR network. Then, Section 4 describes a comparative analysis with state-of-the-art LF SSR methods and validates the effectiveness of the proposed approach through ablation experiments. Finally, a brief summary is provided in Section 5.

2. Related Work

We will introduce LF representation briefly and discuss studies of SSR, ASR and SASR in this section.

2.1. Light Field Representation

As the most important medium for visual perception of the world, light rays carry a wealth of information about the 3D environment. Unlike traditional 2D images, which capture a 2D projection of light through angular integration of the rays in each pixel, LF describes the distribution of light rays in free space and therefore contains richer information. The plenoptic function used to describe the distribution of rays is a 7D function. It depicts the set of rays moving in all directions through each point in 3D space from the perspective of geometrical optics, denoted as $L(x, y, z, \theta, \phi, \gamma, t)$, where (x, y, z) denotes location, (θ, ϕ) denotes angle, γ denotes wavelength and t denotes time. However, these kinds of 7D data are very difficult to capture and process in practical applications. As a result, the LF model was simplified twice in the process of practical application. The first simplification was to remove two dimensions from the plenoptic function, which were the wavelength dimension and time dimension, i.e., reducing the model from 7D to 5D ($L(x, y, z, \theta, \phi)$). The second simplification was to simplify the model into a 4D one by assuming that LF was measured in free space. So, the rays were finally parameterized with the two-plane model $L(u, v, h, w)$, where (u, v) denotes the camera plane and (h, w) denotes the image plane [18].

2.2. Light Field Super-Resolution

SSR refers to the process of reconstructing given LR input images into HR LF images by utilizing complementary information between adjacent SAIs. Performing a single-image SR (SISR) approach for each SAI is a straightforward LF SSR method. But this method cannot exploit the spatial information in a single image and the angular information between different images at the same time. In recent years, many high-efficiency SSR approaches have been created. Wafa et al. [19] introduced a deep learning-based LF SSR network to achieve SR by utilizing full 4D LF angular information and spatial information. Zhang et al. [20] suggested an end-to-end LF SSR network via multiple-epipolar geometry, which divides all views in the LF into four view stacks and feeds them into four different branches to obtain more complete sub-pixel details. Wang et al. [21] studied local feature aggregation and global feature aggregation for LF SSR. Kar et al. [22] suggested adding an adaptive module to a pre-trained SISR framework to improve resolution. Chen et al. [23] used heterogeneous imaging to recover fine textures at large magnification factors for LF SSR. Cheng et al. [24] used a zero-shot learning network for LF SSR through three subtasks. Yao et al. [25] put forward a LF SSR approach via multi-modality fusion, and adaptively enhanced the fused features through a frequency-focusing mechanism. The transformer-based LF SSR method [26] has also achieved excellent results, and this method treats LFSR as a sequence-to-sequence reconstruction task. None of the above methods can maintain the parallax structure of LF images under the premise of fully extracting effective features.

ASR refers to the process of reconstructing dense LF SAIs from a given set of sparse views, i.e., synthesizing new views to improve the LF angular resolution. Densely sampled LF is able to provide not only natural refocus detail, but also smooth parallax offsets, which has led to a wide range of applications for ASR in recent years. Therefore, achieving high-efficiency ASR has become an active research direction. For example, Yun et al. [27] proposed geometrically aware LF angular SR based on a multi-receiving field network, which is very simple and consists of only two main modules. Liu et al. [28] proposed an efficient LF angular SR based on sub-aperture feature learning and macro pixel upsampling, which can explore the spatial-angular correlation well. Jin et al. [29] proposed deep coarse-to-fine dense LF reconstruction based on flexible sampling and geometric perception fusion.

SASR refers to the simultaneous execution of SSR and ASR on LF images, which not only improves the spatial resolution of each SAI, but also increases the number of views of LF images. In recent years, there has been an increasing number of studies on SASR. Zhou et al. [30] proposed end-to-end LF SASR based on a parallax structure reservation strategy, which can generate an HR dense LF from an LR sparse LF with detailed textures. Duong et al. [31] and Sheng et al. [32] also suggested approaches that can obtain high-quality SASR reconstruction results.

In recent years, with the rapid development of deep convolutional neural networks (CNNs), whether it is SSR, ASR or SASR, most people used learning-based methods. These methods can not only save time to a large extent, but also accurately restore high-frequency details of LF reconstructed images, which lays a good foundation for subsequent research.

3. Proposed Method

In this section, the networks proposed in this article will be described in detail. As mentioned in Section 2.1 above, LF uses a two-plane model $L(u, v, h, w)$ to parameterize the rays. Therefore, a 4D LR LF image can be represented as $I^{lr} \in R^{U \times V \times H \times W}$, where $U \times V$ represents angular resolution and $H \times W$ represents spatial resolution. In this paper, we use a square SAI array (i.e., $U = A = V$), so a 4D LF image contains an $A \times A$ array of SAIs. Our goal is to obtain an HR LF image $I^{hr} \in R^{U \times V \times \alpha H \times \alpha W}$ from its LR counterpart I^{lr} , and α denotes the spatial magnification factor. In order to decrease computing complexity, the proposed approach is only executed on the Y channel of the LF image. A bicubic interpolation algorithm is performed to obtain the super-resolution of the Cr and Cb channels, and after that, the Y, Cr and Cb channels are converted into an RGB image.

3.1. Overall Network Framework

The overall network framework of our method is illustrated in Figure 1, and is composed of three major components: a feature extraction block (FEB), view interaction blocks and a reconstruction block. Firstly, the LR SAI $S_i^{lr} \in R^{1 \times H \times W}$ ($i = 1, 2, \dots, A^2$) is fed into the reshaping block of the upper branch, which is connected along the channel dimension to obtain $S^{lr} \in R^{A^2 \times H \times W}$; then, the FEB is used to extract the global-view feature $F^g \in R^{C \times H \times W}$ from S^{lr} . Different from the upper branch, the lower branch does not need to perform reshaping operation, but directly extracts features, and maps each SAI to the depth feature representation to obtain the hierarchical feature $F_i^l \in R^{C \times H \times W}$ ($i = 1, 2, \dots, A^2$). For different views, the weights in the FEB of the lower branch are shared. In order to thoroughly investigate the correlation between all views, the extracted features are subsequently passed continuously through four double-branched view interaction blocks. This process enhances and connects the features in a sequential manner, effectively maintaining the parallax structure. The view interaction block is composed of four parts: an inter-view unit (InterU), a global-view feature, an intra-view unit (IntraU) and a multi-view aggregation block (MAB). These are collectively referred to as IGIM. Finally, the HR SAIs $S_i^{hr} \in R^{1 \times \alpha H \times \alpha W}$ ($i = 1, 2, \dots, A^2$) are obtained through the reconstruction block, which is composed of a feature fusion block and an upsampling block.

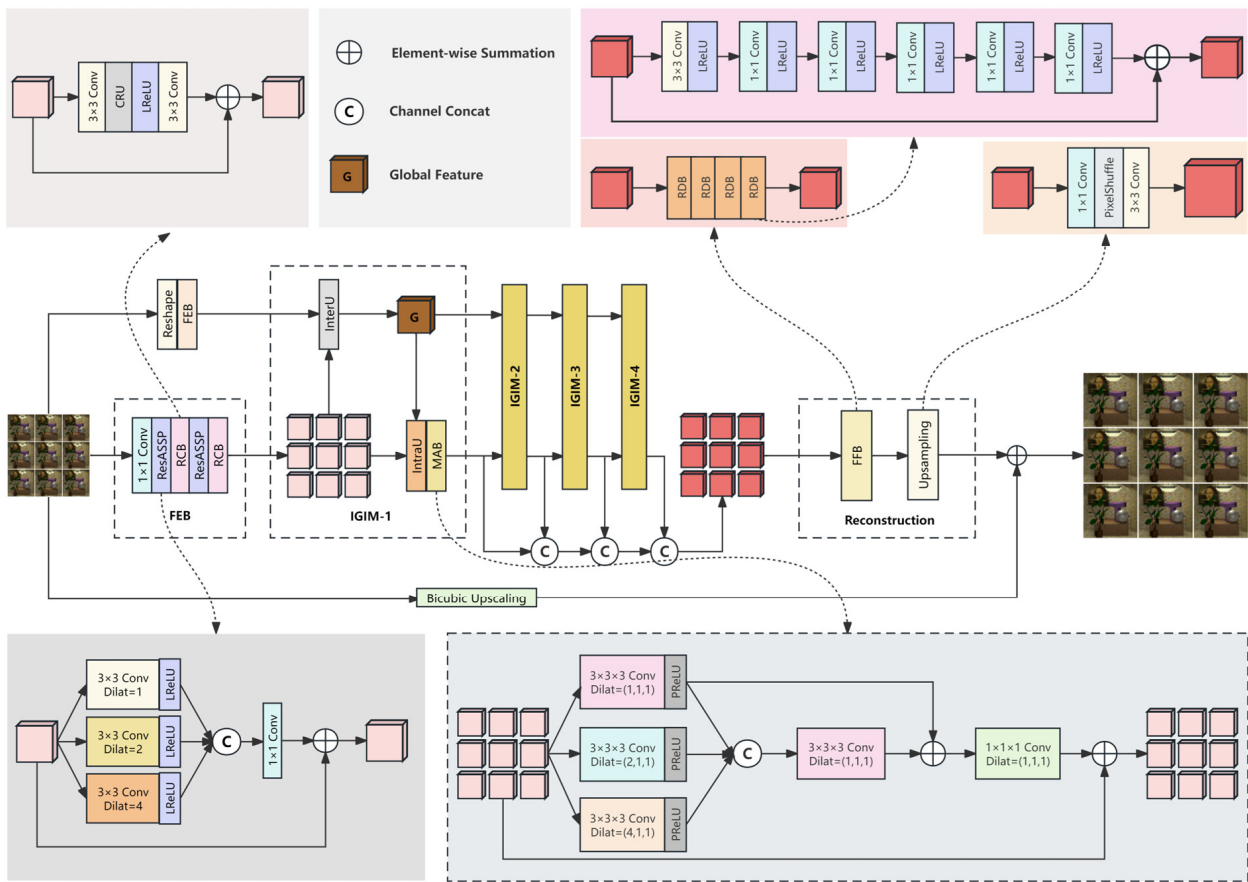


Figure 1. The overall network framework of the proposed LF-IGIM.

3.2. Feature Extraction Block (FEB)

Our FEB is treated separately for each input SAI, and its weights are shared between SAIs. Firstly, a 1×1 convolution in the FEB is introduced to extract the initial features of a single SAI and then feed them into cascaded ResASPP and a residual channel-reconstruction block (RCB) for further feature extraction. The depth C is set to 32. The architecture of ResASPP is shown in Figure 1, which is to extend receptive field to gain rich contextual information. We can extract multi-scale information by combining 3×3 dilated convolutions with three different expansion rates in parallel, which are 1, 2 and 4. The Leaky ReLU (LReLU) is an activation function, and its leaky factor is 0.1. Finally, the features obtained by the three parallel branches are connected and then fused through 1×1 convolution. In addition, we design an RCB to successfully reduce the channel redundancy between features by introducing a channel reconstruction unit (CRU) [33]. As shown in Figure 1, our RCB consists of two 3×3 convolutions, a CRU and a LReLU activation function.

The CRU is constructed by using a split-transform-fuse approach to reduce channel redundancy, as shown in Figure 2. Firstly, the aim of the splitting part is to divide the input features $F \in R^{c \times h \times w}$ with C channels into the upper-branch features $F_{up} \in R^{\frac{\alpha C}{r} \times h \times w}$ with αC channels and the lower-branch features $F_{low} \in R^{\frac{(1-\alpha)C}{r} \times h \times w}$ with $(1-\alpha)C$ channels, where α is the split ratio, the numerical range is between 0~1, and r is the squeeze ratio. The channels of those feature maps are compressed by a 1×1 convolution to improve the calculation speed.

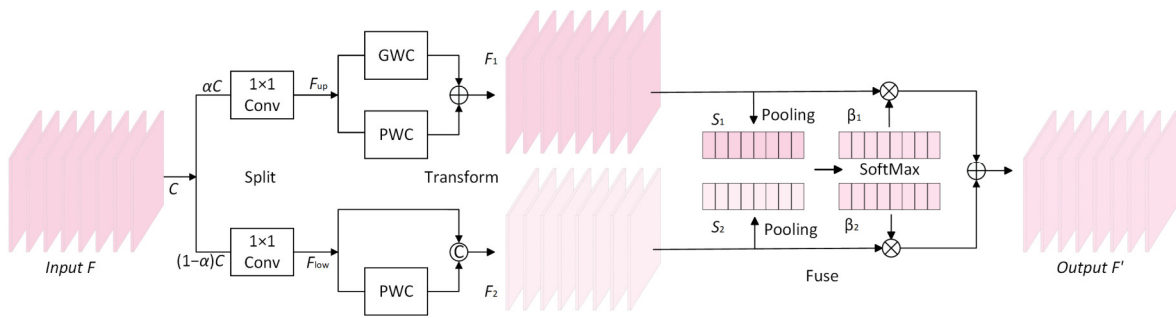


Figure 2. The architecture of the CRU.

Secondly, the aim of the transforming part is to extract the rich representative feature maps $F_1 \in R^{c \times h \times w}$ in the upper-branch features F_{up} through efficient group-wise convolution (GWC) and point-wise convolution (PWC). GWC, which was first introduced in AlexNet, can be regarded as a sparse convolution connection method in which each output channel is linked to only a certain group of input channels. In addition, PWC is used to keep the information flowing across channels and enable dimensionality reduction by reducing the number of filters. The process can be represented as

$$F_1 = M^G F_{up} + M^{P_1} F_{up} \tag{1}$$

where $M^G \in R^{\frac{ac}{g^r} \times k \times k \times c}$ and $M^{P_1} \in R^{\frac{ac}{r} \times 1 \times 1 \times c}$ are the learnable weight matrices of GWC and PWC, $k \times k$ is the size of convolution kernel and g is the group size in the experiment ($g = r = 2$). At the same time, only PWC is used to acquire the feature maps $F_2 \in R^{c \times h \times w}$ with shallow hidden details in the lower-branch features F_{low} as a supplement to the upper-branch feature mapping, which can be represented as

$$F_2 = M^{P_2} F_{low} \cup F_{low} \tag{2}$$

where $M^{P_2} \in R^{\frac{(1-a)c}{r} \times 1 \times 1 \times (1-\frac{1-a}{r})c}$ is also a weight matrix of PWC, and \cup means concatenation.

Finally, the aim of the fusing part is to obtain the global spatial information through the global average pooling operation to obtain $S_m \in R^{c \times 1 \times 1}$. The process is as follows:

$$S_m = Pooling(F_m) = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W F_c(i, j), m = 1, 2 \tag{3}$$

Then, the upper-branch and lower-branch global channel-weight descriptors are superimposed together, and a channel-weight soft attention operation is designed to create the important feature vectors β_1 and β_2 as follows:

$$\beta_1 = \frac{e^{s_1}}{e^{s_1} + e^{s_2}}, \beta_2 = \frac{e^{s_2}}{e^{s_1} + e^{s_2}}, \beta_1 + \beta_2 = 1 \tag{4}$$

Under the guidance of these two vectors, the upper-branch feature maps and the lower-branch feature maps are combined in a channel-wise manner to obtain output features $F_l \in R^{c \times h \times w}$ with reduced redundancy in the channel dimension, and the procedure is as follows:

$$F_l = \beta_1 F_1 + \beta_2 F_2 \tag{5}$$

From the above-detailed process, it can be seen that after integrating the channel reconstruction unit (CRU) into the channel-reconstruction block (RCB), we can not only reduce the redundant calculation of feature extraction blocks in the channel dimension of feature mapping, but also promote the learning of representative features in the feature extraction process. This is an indispensable way to efficiently acquire important features.

3.3. View Interaction Block

The view interaction block is composed of an inter-view unit (InterU), a global-view feature, an intra-view unit (IntraU) and a multi-view aggregation block (MAB), which is an improvement of intra-inter-view feature interaction [34]. Specifically, we can reduce the redundant calculation of feature mapping in the spatial dimension after integrating a spatial reconstruction unit (SRU) into InterU and IntraU in the view interaction block, so as to promote the efficient use of complementary information between all views in the interaction process. Furthermore, our MAB uses a smaller number of 3D convolutions to efficiently model the correlation between all SAIs to further achieve SSR. The architecture of InterU is shown in Figure 3a and the architecture of IntraU is shown in Figure 3b; the former uses the hierarchical feature F_i^l obtained by the lower branch to update the global inter-view feature F^g , and the latter uses the global inter-view feature F^g obtained by the upper branch to update the hierarchical feature F_i^l .

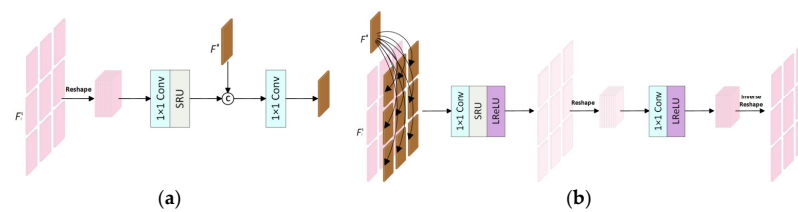


Figure 3. The architecture of two important components in the view interaction block: (a) InterU and (b) IntraU.

The SRU in the structure is constructed by using a separate-reconstruct method to reduce spatial redundancy, as depicted in Figure 4. Firstly, the aim of the separating part is to subdivide the feature maps with a large amount of information and the feature maps with a small amount of information corresponding to the spatial content in the input feature $X \in R^{c \times h \times w}$, and it first uses group normalization to standardize X , which can be portrayed as follows:

$$X_0 = GN(X) = \gamma \frac{X - \mu}{\sqrt{\sigma^2 + \epsilon}} + \beta \tag{6}$$

where σ and μ are the standard and mean deviation in X ; γ and β are trainable affine transformations; and ϵ is a small positive constant. Each normalized correlation weight $W_\gamma \in R^c$ is determined to represent the importance of each feature map, which can be obtained from the following formula:

$$W_\gamma = \{w_i\} = \frac{\gamma_i}{\sum_{j=1}^c \gamma_j}, i, j = 1, 2, \dots, c \tag{7}$$

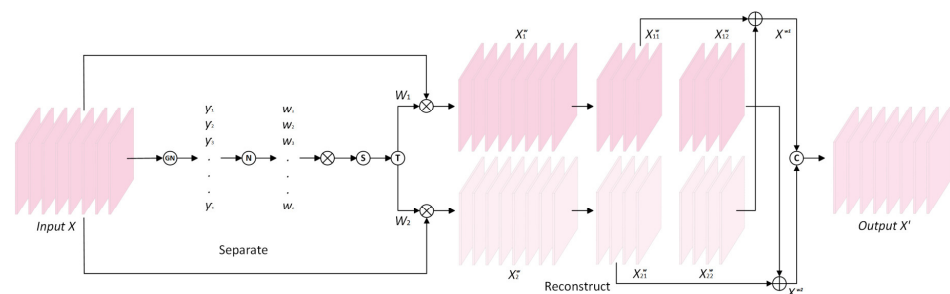


Figure 4. The architecture of the SRU.

Then, W_γ is mapped to 0~1 through the sigmoid function, and the threshold is set for gating. The weight exceeding the threshold is reset to 1 to obtain the information weight

W_1 ; otherwise it is set to 0, and the non-information weight W_2 is obtained. The entire process of obtaining W can be depicted as

$$W = Gate(Sigmoid(W_\gamma(GN(X)))) \quad (8)$$

Finally, W_1 and W_2 are multiplied by the input features X to create weighted features X_1^w with a large amount of information and weighted features X_2^w with a small amount of information.

The aim of the reconstructing part is to first cross-add the informative features X_1^w and the less informative features X_2^w , fully combine the two pieces of information with different weights, and then connect them to obtain the output features $X' \in R^{c \times h \times w}$ with reduced redundancy in the spatial dimension. The specific process is as follows:

$$\begin{cases} X_1^w = W_1 \otimes X, \\ X_2^w = W_2 \otimes X, \\ X_{11}^w \oplus X_{22}^w = X^{w1}, \\ X_{12}^w \oplus X_{21}^w = X^{w2}, \\ X^{w1} \cup X^{w2} = X'. \end{cases} \quad (9)$$

where \oplus is element-wise summation, \otimes is element-wise multiplication and \cup is concatenation.

In addition, the multi-view aggregation block (MAB) in the view interaction block is based on 3D convolution to model the correlation between all SAIs, as shown in Figure 1. In the MAB, the hierarchical features $F_i^l \in R^{C \times H \times W}$ ($i = 1, 2, \dots, A^2$) obtained by the lower branch are firstly stacked along the angular dimension to obtain the feature $\tilde{F}^l \in R^{C \times H \times W \times A^2}$, which is processed by three $3 \times 3 \times 3$ convolutions. Then, the features obtained by the combination of three parallel branches are connected. The features are fused by using a $3 \times 3 \times 3$ convolution with a dilation rate of (1, 1, 1). A $1 \times 1 \times 1$ convolution is used to process the fused features, which are added to the input features to obtain the output hierarchical features. The activation function PReLU in the MAB is an improvement of LReLU with a leaky factor of 0.02.

3.4. Reconstruction Block

As shown in Figure 1, the reconstruction block includes a feature fusion block (FFB) and an upsampling block. Firstly, the FFB is composed of four cascaded residual dense blocks (RDBs), and each RDB is composed of 6 convolutional layers and 6 activation functions. Our FFB can fuse shallow features and deep features generated by view interaction blocks to obtain high reconstruction accuracy. In each RDB, the input features are performed by a 3×3 convolution; then, the following dense connected layers are processed by 1×1 convolutions. The output F^i of the i -th ($2 \leq i \leq 5$) convolution of each RDB can be portrayed as

$$F^i = \sigma_l \left(C_i \left(\left[F^1, \dots, F^{i-1} \right] \right) \right) \quad (10)$$

where σ_l means the activation function LReLU, C_i represents the i -th convolution and $[\cdot]$ denotes concatenation. Then, the features are fused by the last 1×1 convolution, and the local residual connection is eventually designed. In the RDB, the numbers of filters are $4C$, C , C , C , C , and $4C$ from left to right. The fused features are then fed to the upsampling block, which consists of a 1×1 convolution, a 3×3 convolution and a PixelShuffle layer between them. The LR feature maps can be converted into HR feature maps through the upsampling block. Finally, the features generated by the upsampling block are added to the initial features to achieve global residual learning to an HR LF.

3.5. Loss Function

The loss function used in our training process is the absolute value loss function (L_1). The loss function of a batch of training pairs $\{I_k^{HR}, I_k^{GT}\}_{k=1}^n$ containing n output HR

LF images of the SR network and their corresponding ground-truth (GT) LF images is calculated as follows:

$$L_1 = \frac{1}{n} \sum_{k=1}^n \left\| I_k^{HR} - I_k^{GT} \right\|_1 \quad (11)$$

4. Experiments

4.1. Datasets and Implementation Details

The five public datasets used in this paper are EPFL [35], HCInew [36], HCIold [37], INRIA [38] and STFGantry [39]. The last one is a large-disparity LF dataset. Table 1 shows the particulars of the scene categories, which include the number of training scenes and test scenes for each dataset. The angular resolution of all input LF images is 5×5 and the spatial resolution is 32×32 . Only SSR is applied, with upscaling factors of 2 and 4. The peak signal-to-noise ratio (PSNR) and structural similarity index measure (SSIM) of the Y channel of the images are calculated as experimental evaluation indicators.

Table 1. Particulars of five datasets used in our experiments.

| Datasets | Training # | Testing # | Scene |
|----------------|------------|-----------|------------|
| EPFL [35] | 70 | 10 | Real-world |
| HCInew [36] | 20 | 4 | Synthetic |
| HCIold [37] | 10 | 2 | Synthetic |
| INRIA [38] | 35 | 5 | Real-world |
| STFGantry [39] | 9 | 2 | Real-world |

Our LF-IGIM was implemented in PyTorch and trained with an NVIDIA GeForce RTX 3050 Laptop GPU. We used Adam optimizer to optimize our network. Our training data were augmented by randomly flipping and rotating the images by 90° . During the training process, the following parameters remain unchanged for both the $2 \times$ SR network and the $4 \times$ SR network: the batch size was set to 4, the number of channels C was set to 32 and the initial learning rate was set to 2×10^{-4} , which was decreased by multiplying by a coefficient of 0.5 for every 15 epochs. The total number of training epochs was set to 50.

4.2. Comparison with State-of-the-Art Methods

Our LF-IGIM network proposed in this paper was compared with 8 other LF SSR methods, namely LFBM5D [7], GB [8], resLF [12], LF-ATO [11], LF-InterNet [14], LF-DFnet [15], DPT [26] and LF-IINet [34]. Among them, LFBM5D [7] and GB [8] are optimization-based LF SSR methods, and the last six methods are learning-based LF SSR methods. The method proposed in this paper is also a learning-based LF SSR method. For a fair comparison, all of these methods were performed by using the same training scenes.

4.2.1. Quantitative Results

The results of the proposed LF-IGIM were quantitatively compared with the other eight approaches mentioned above, and the PSNR/SSIM values (between the reconstructed and GT images) obtained from the test scenes are shown in Tables 2 and 3. Specifically, Table 2 shows the PSNR [dB]/SSIM scores compared with the other eight SOTA methods for $2 \times$ SR. Table 3 shows the PSNR [dB]/SSIM scores compared with the other eight SOTA methods for $4 \times$ SR. The best scores are highlighted in bold, and the second best scores are underlined. If the above methods involve a method that does not disclose the complete code, the numerical results in the paper are directly cited. It is worth mentioning that the PSNR/SSIM values obtained on each test dataset here were calculated first for each pair of SAIs for each scene, and then, we calculated the average PSNR/SSIM values for all SAIs. However, the averages in the table refer to the average scores for all test scenes in the five datasets.

Table 2. PSNR [dB]/SSIM scores compared with the other 8 SOTA methods for 2×SR.

| Method | EPFL | HCInew | HCInold | INRIA | STFGantry | Average |
|------------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|
| Bicubic | 29.50/0.9350 | 31.69/0.9335 | 37.46/0.9776 | 31.10/0.9563 | 30.82/0.9473 | 32.11/0.9499 |
| LFBM5D [7] | 31.15/0.9545 | 33.72/0.9548 | 39.62/0.9854 | 32.85/0.9659 | 33.55/0.9718 | 34.18/0.9665 |
| GB [8] | 31.22/0.9591 | 35.25/0.9692 | 40.21/0.9879 | 32.76/0.9724 | 35.44/0.9835 | 34.98/0.9744 |
| resLF [12] | 32.75/0.9672 | 36.07/0.9715 | 42.61/0.9922 | 34.57/0.9784 | 36.89/0.9873 | 36.58/0.9793 |
| LF-ATO [11] | 34.22/0.9752 | 37.13/0.9761 | 44.03/0.9940 | 36.16/0.9841 | 39.20/0.9922 | 38.15/0.9843 |
| LF-InterNet [14] | 34.14/0.9761 | 37.28/0.9769 | 44.45/0.9945 | 35.80/0.9846 | 38.72/0.9916 | 38.08/0.9847 |
| LF-DFnet [15] | 34.44/0.9766 | 37.44/0.9786 | 44.23/0.9943 | 36.36/0.9841 | 39.61/0.9935 | 38.42/0.9854 |
| DPT [26] | 34.48/0.9759 | 37.35/0.9770 | 44.31/0.9943 | 36.40/0.9843 | 39.52/0.9928 | 38.41/0.9849 |
| LF-IINet [34] | <u>34.68/0.9771</u> | <u>37.74/0.9789</u> | <u>44.84/0.9948</u> | <u>36.57/0.9853</u> | <u>39.86/0.9935</u> | <u>38.74/0.9859</u> |
| LF-IGIM (Ours) | 34.85/0.9777 | 38.02/0.9799 | 44.86/0.9949 | 36.82/0.9857 | 40.42/0.9942 | 38.99/0.9865 |

The best results are highlighted in bold and the second best results are underlined.

Table 3. PSNR [dB]/SSIM scores compared with other 8 SOTA methods for 4×SR.

| Method | EPFL | HCInew | HCInold | INRIA | STFGantry | Average |
|------------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|
| Bicubic | 25.14/0.8311 | 27.61/0.8507 | 32.42/0.9335 | 26.82/0.8860 | 25.93/0.8431 | 27.58/0.8689 |
| LFBM5D [7] | 26.61/0.8689 | 29.13/0.8823 | 34.23/0.9510 | 28.49/0.9137 | 28.30/0.9002 | 29.35/0.9032 |
| GB [8] | 26.02/0.8628 | 28.92/0.8842 | 33.74/0.9497 | 27.73/0.9085 | 28.11/0.9014 | 28.90/0.9013 |
| resLF [12] | 27.46/0.8507 | 29.92/0.9011 | 36.12/0.9651 | 29.64/0.9339 | 28.99/0.9214 | 30.43/0.9144 |
| LF-ATO [11] | 28.64/0.9130 | 30.97/0.9150 | 37.06/0.9703 | 30.79/0.9490 | 30.79/0.9448 | 31.65/0.9384 |
| LF-InterNet [14] | 28.67/0.9143 | 30.98/0.9165 | 37.11/0.9715 | 30.64/0.9486 | 30.53/0.9426 | 31.59/0.9387 |
| LF-DFnet [15] | 28.77/0.9165 | 31.23/0.9196 | 37.32/0.9718 | 30.83/0.9503 | 31.15/0.9494 | 31.86/0.9415 |
| DPT [26] | 28.93/0.9167 | 31.19/0.9186 | 37.39/0.9720 | 30.96/0.9502 | 31.14/0.9487 | 31.92/0.9412 |
| LF-IINet [34] | <u>29.11/0.9194</u> | <u>31.36/0.9211</u> | <u>37.62/0.9737</u> | <u>31.08/0.9516</u> | <u>31.21/0.9495</u> | <u>32.08/0.9431</u> |
| LF-IGIM (Ours) | 29.12/0.9208 | 31.50/0.9231 | 37.70/0.9739 | 31.15/0.9524 | 31.43/0.9520 | 32.18/0.9444 |

The best results are highlighted in bold and the second best results are underlined.

In Table 2, we can see that our LF-IGIM obtains the best scores on each test dataset. Compared with the two traditional SSR methods (LFBM5D and GB), our LF-IGIM obtains much higher scores because of the robust representation capability of deep CNNs. The PSNR value obtained by our LF-IGIM on the large-disparity dataset (STFGantry) is 0.56 dB higher than that of the second-ranked LF-IINet. The average PSNR and SSIM values are 0.25 dB higher and 0.006 higher than LF-IINet.

In Table 3, we can see that our LF-IGIM achieves the best scores on each dataset as well. Compared with the two traditional SSR methods (LFBM5D and GB), our LF-IGIM achieves 2.83 dB and 3.28 dB gains on average for 4×SR. Compared with the second-ranked LF-IINet, the SSIM value obtained by our LF-IGIM on the small-disparity dataset (HCInew) is increased by 0.0020. In addition, compared with LF-IINet, the average PSNR value obtained by our method is increased by 0.10 dB, and the average SSIM value is also increased by 0.0013. This indicates that our LF-IGIM fully utilizes the complementary information between all SAIs.

4.2.2. Qualitative Results

The results of our LF-IGIM were qualitatively compared with the eight methods mentioned above. The 2×SR visualization results obtained by our LF-IGIM and the other eight methods on the test scene *Lego Knights* in the large-disparity dataset (STFGantry) are shown in Figure 5a, and the 2×SR visualization results obtained on the test scene *herbs* in the small-disparity dataset (HCInew) are shown in Figure 5b. The enlarged patches are highlighted by using red box and blue box, and the green arrow is utilized to highlight the comparison. In Figure 5a, the result obtained by our method is clearer and closer to the GT than the other methods in the small circles in the red box. The result obtained by our method is also clearer in the blue box than the other methods. The reliability of our LF-IGIM on large-disparity datasets is demonstrated. In Figure 5b, the textures we obtain in both

boxes are sharper and closer to the GT. Our LF-IGIM presents better fine-grained details. The $4\times$ SR visualization results obtained by our method and the other eight methods on the test scenes *origami* and *bicycle* in HCInew are shown in Figures 6a and 6b, respectively. In Figure 6a, our method obtains a yellow grid in the red box that is clearer and closer to the ground-truth than other methods. In Figure 6b, the results of our method on the second-row books are closer to GT than the others, especially the outline of the gray book in the red box. The outline of greenery in the lower left corner of the blue box is significantly closer to the GT than the other methods. The reliability of our LF-IGIM on small-disparity datasets is demonstrated.

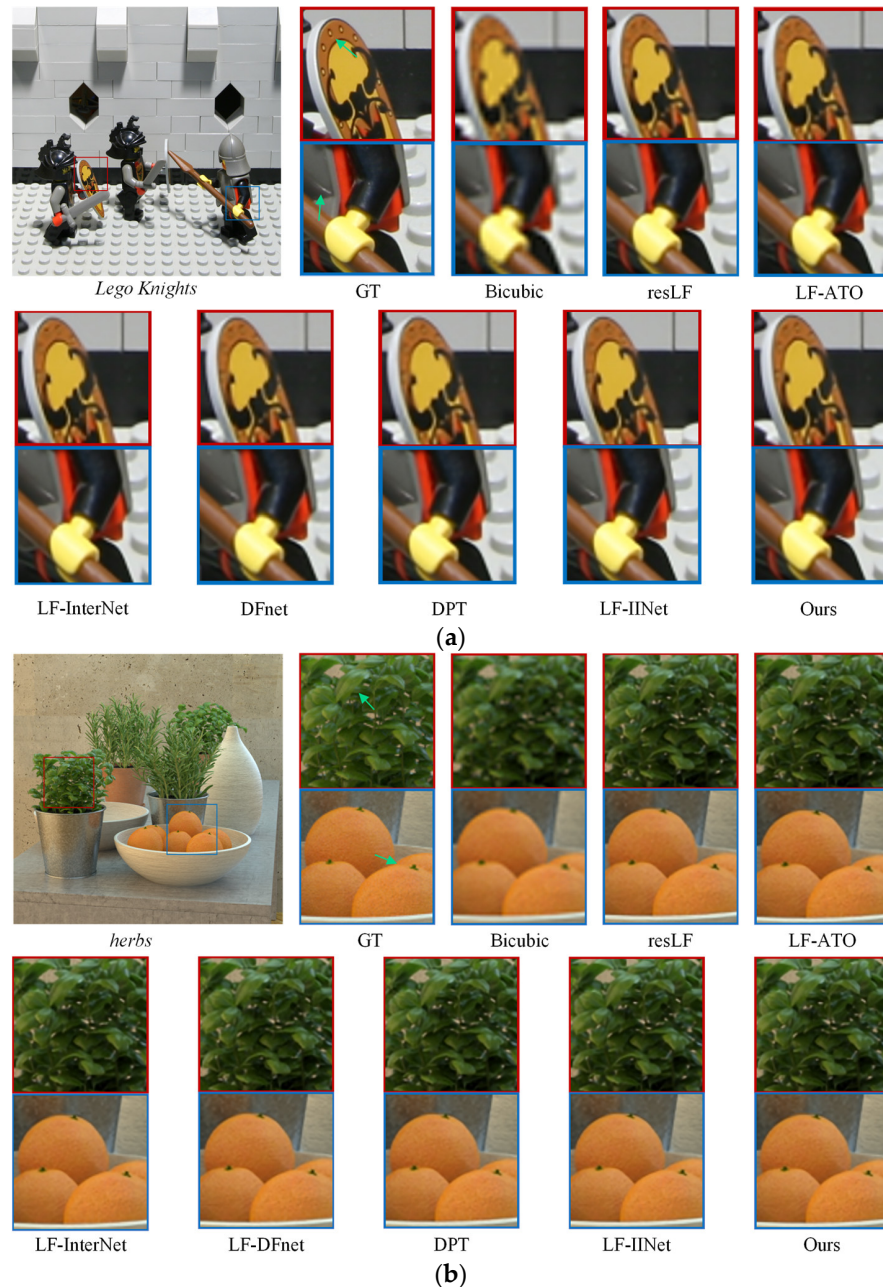


Figure 5. The $2\times$ SR visualization results of different methods on two test scenes: (a) *Lego Knights* and (b) *herbs*.



Figure 6. The 4×SR visualization results of different methods on two test scenes: (a) *origami* and (b) *bicycle*.

We also provide the 4×SR results of the EPIs (highlighted in the green box) for qualitative comparison, as shown in Figure 7. The EPIs were constructed by concatenating the same line from all LF SAIs along the height dimension. The line structure in the EPIs shows the parallax structure of the reconstructed LF. For the two scenes, it is obvious that the EPI result of Bicubic is very blurry. This means that the method cannot maintain LF parallax consistency. Our LF-IGIM obtains the most fine-grained lines in the EPIs, which proves the effectiveness of the parallax structure reservation performed by our network.

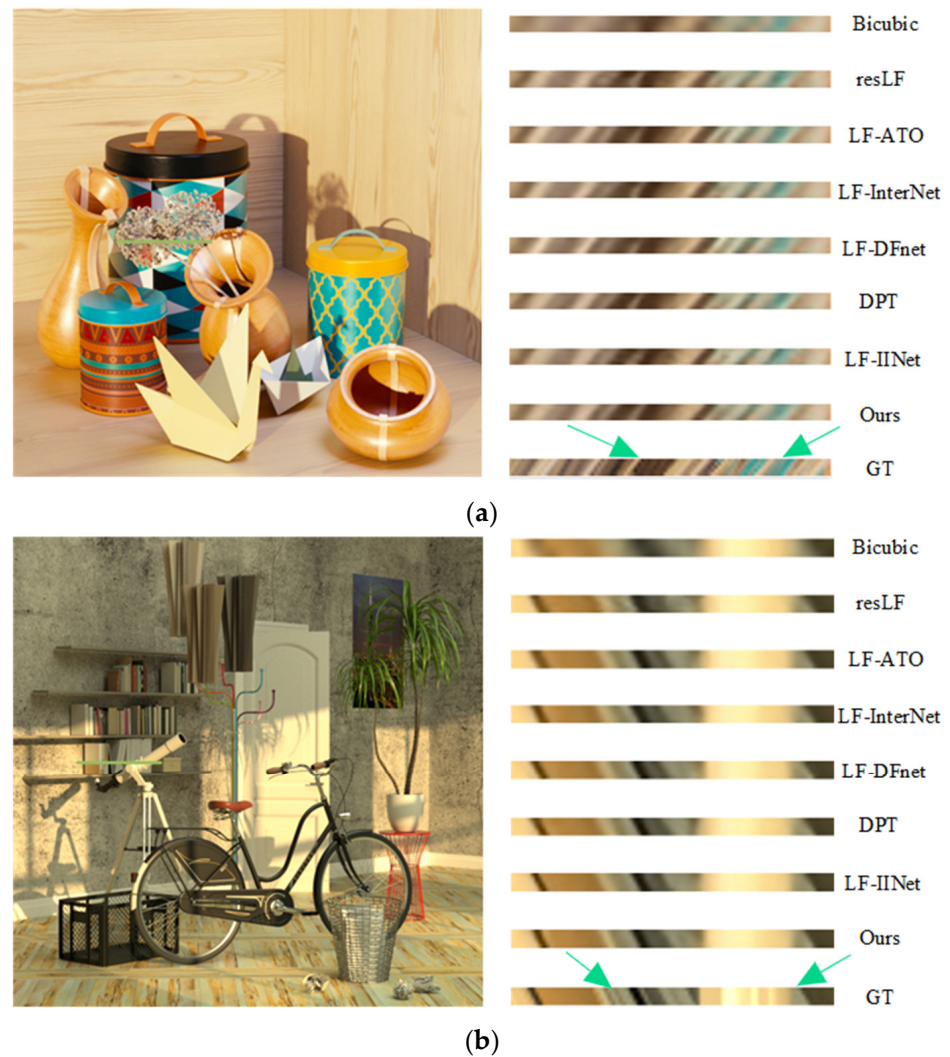


Figure 7. The $4\times$ SR results of the EPIs obtained by different methods on two test scenes: (a) *origami* and (b) *bicycle*.

4.2.3. Parameters and FLOPs

We provide the number of parameters and floating point operations (#Params and FLOPs) of different networks, and the FLOPs were measured with an input LF image size of $5 \times 5 \times 32 \times 32$. The specific results are shown in Tables 4 and 5.

Table 4. Specific results of #Params, FLOPs and average PSNR [dB]/SSIM scores for $2\times$ SR.

| Method | #Params. | FLOPs | Ave. PSNR [dB]/SSIM |
|------------------|----------|----------|---------------------|
| resLF [12] | 6.35 M | 37.06 G | 36.58/0.9793 |
| LF-ATO [11] | 1.51 M | 597.66 G | 38.15/0.9843 |
| LF-InterNet [14] | 4.80 M | 47.46 G | 38.08/0.9847 |
| LF-DFnet [15] | 3.94 M | 57.22 G | 38.42/0.9854 |
| DPT [26] | 3.73 M | 57.44 G | 38.41/0.9849 |
| LF-IINet [34] | 4.84 M | 56.16 G | 38.74/0.9859 |
| LF-IGIM (ours) | 5.85 M | 65.91 G | 38.99/0.9865 |

The best average PSNR [dB]/SSIM scores are highlighted in bold.

Table 5. Specific results of #Params, FLOPs and average PSNR [dB]/SSIM scores for 4×SR.

| Method | #Params. | #FLOPs | Aver. PSNR [dB]/SSIM |
|------------------|----------|----------|----------------------|
| resLF [12] | 6.79 M | 39.70 G | 30.43/0.9144 |
| LF-ATO [11] | 1.66 M | 686.99 G | 31.65/0.9384 |
| LF-InterNet [14] | 5.23 M | 50.10 G | 31.59/0.9387 |
| LF-DFnet [15] | 3.99 M | 57.31 G | 31.86/0.9415 |
| DPT [26] | 3.78 M | 58.64 G | 31.92/0.9412 |
| LF-IINet [34] | 4.89 M | 57.42 G | 32.08/0.9431 |
| LF-IGIM (ours) | 5.90 M | 67.25 G | 32.18/0.9444 |

The best average PSNR [dB]/SSIM scores are highlighted in bold.

In Table 4, it is obvious that the FLOP of our LF-IGIM is apparently lower than that of LF-ATO. This is because LF-ATO exploits all SAIs to obtain one HR SAI, which needs a high FLOP to obtain all HR SAIs in an end-to-end manner. Compared with LF-InterNet, the #Param and FLOP of our LF-IGIM increase by 1.05 M and 18.45 G, respectively, while the average PSNR and average SSIM values increase by 0.91 dB and 0.0018, respectively. Compared with the second-best LF-IINet, the number of parameters of our LF-IGIM is only increased by 1.01 M, and the number of FLOPs is only increased by 9.75 G, but the average PSNR value and average SSIM value are increased by 0.25 dB and 0.0006, respectively. Because we used an FFB cascaded by RDBs to efficiently fuse the hierarchical features to keep the consistency of the LF parallax structure, from our ablation experiment, it can be seen that after reducing the number of RDBs, the number of parameters of the network is significantly reduced by 1.15 M. This is because the RDB is densely connected, and the number of parameters is large. In Table 5, compared with LF-ATO, although the number of parameters of our LF-IGIM increases by 4.24 M, the number of FLOPs decreases by 619.74 G, and the average PSNR value and average SSIM value increase by 0.53 dB and 0.0060, respectively. Compared with LF-InterNet, the numbers of parameters and FLOPs of our LF-IGIM increase by 0.67 M and 17.15 G, respectively, while the average PSNR value and average SSIM value increase by 0.59 dB and 0.0057, respectively. In summary, our LF-IGIM achieves state-of-the-art SR performance with reasonable model size and computational cost.

4.3. Ablation Study and Discussion

The proposed LF-IGIM was compared with three adjusted network structures to verify the effectiveness of the proposed RCB, SRU and RDB. The quantitative results of the ablation study are shown in Table 6, and the best results are highlighted in bold. It is important to note that the study of the ablation experiment was performed with the amplification factor set to 4. The five datasets used were consistent with the previously described datasets, and the parameters remained unchanged. The three adjusted network structures were as follows: (1) *LF-IGIM w/o RCB*: the network was obtained by directly removing the RCB in the FEB; (2) *LF-IGIM w/o SRU*: the network was obtained by directly removing the SRU in InterU and IntraU; (3) *LF-IGIM w/o RDB*: the network was based on the change in the number of RDBs in the FFB from four to one.

Table 6. The PSNR [dB]/SSIM values of the ablation study for 4×SR.

| Network | #Params | EPFL | HCInew | HCIold | INRIA | STFgantry | Average |
|------------------------|---------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|
| <i>LF-IGIM w/o RCB</i> | 5.82 M | 28.93/0.9194 | 31.39/0.9216 | 37.59/0.9735 | 30.96/0.9517 | 31.40/0.9514 | 32.05/0.9435 |
| <i>LF-IGIM w/o SRU</i> | 5.24 M | 29.09/0.9201 | 31.46/0.9222 | 37.65/0.9736 | 31.04/0.9521 | 31.40/0.9514 | 32.13/0.9439 |
| <i>LF-IGIM w/o RDB</i> | 4.75 M | 28.85/0.9164 | 31.17/0.9188 | 37.29/0.9718 | 30.81/0.9495 | 30.86/0.9461 | 31.80/0.9405 |
| LF-IGIM | 5.90 M | 29.12/0.9208 | 31.50/0.9231 | 37.70/0.9739 | 31.15/0.9524 | 31.43/0.9520 | 32.18/0.9444 |

The best results are highlighted in bold.

Firstly, Table 6 shows that the number of parameters is reduced by only 0.08 M when the RCB is removed from the LF-IGIM. However, the PSNR values obtained on both EPFL

and INRIA are reduced by 0.19 dB, and the final average PSNR value is also reduced by 0.13 dB. Moreover, the SSIM value obtained by *LF-IGIM w/o RCB* on HCInew is also reduced by 0.0015 and the final average SSIM value is also reduced by 0.0009 compared to the full LF-IGIM model. Therefore, the proposed RCB not only has a small number of parameters, but also successfully reduces the channel redundancy in the feature extraction process, effectively extracts useful information and improves SR performance. Secondly, Table 6 shows that the number of parameters is reduced by only 0.66 M when the SRU is removed from the LF-IGIM. However, the PSNR value obtained on the dataset INRIA is reduced by 0.11 dB, and the final average PSNR value is also reduced by 0.05 dB. Moreover, the SSIM value obtained by the *LF-IGIM w/o SRU* on the dataset HCInew is also reduced by 0.0009 and the final average SSIM value is also reduced by 0.0005 compared to the full LF-IGIM model. It can be seen that the SRU not only has a small number of parameters, but also successfully reduces the spatial redundancy in the process of feature enhancement, and effectively learns the complementary information between all views, which improves the SR performance. Finally, Table 6 shows that the number of parameters is also reduced by 1.15 M when the number of RDBs is reduced from four to one. The PSNR value obtained on the dataset STFgantry is reduced by 0.57 dB, and the final average PSNR value is also reduced by 0.38 dB. Moreover, compared with the full LF-IGIM model, the SSIM value obtained by *LF-IGIM w/o RDB* on STFgantry is also reduced by 0.0059, and the final average SSIM value is also reduced by 0.0039. It can be seen that although the number of parameters of the RDB is not small, the hierarchical features are efficiently fused during the reconstruction process, and the SR performance is successfully improved.

From the results of the ablation experiment, it can be seen that when we directly remove the RCB from LF-IGIM, the average PSNR value is 0.03 dB lower than that obtained by IINet, but the average SSIM value is 0.0004 higher than it. When we delete the SRU in LF-IGIM, the average PSNR/SSIM value is 0.05 dB/0.0008 higher than that of IINet. But at the same time, we have a 0.35 M higher number of parameters than IINet. In addition, after reducing the number of RDBs, the number of parameters obtained is 0.14M fewer than that of IINet, but the average PSNR/SSIM is also reduced by 0.28 dB/0.0026. Because of the dense connection of our FFB, the number of parameters is large. These are the limitations of our approach. Overall, the submodules we introduced in LF-IGIM improve LF SRR's performance.

In addition, a visualization of the ablation study on the test scene *Tarot Card S* in the dataset Stanford_Gantry is shown in Figure 8. We can see from the visualization results obtained by removing the SRU, RCB and RDB from LF-IGIM that the lines and letters are blurred.

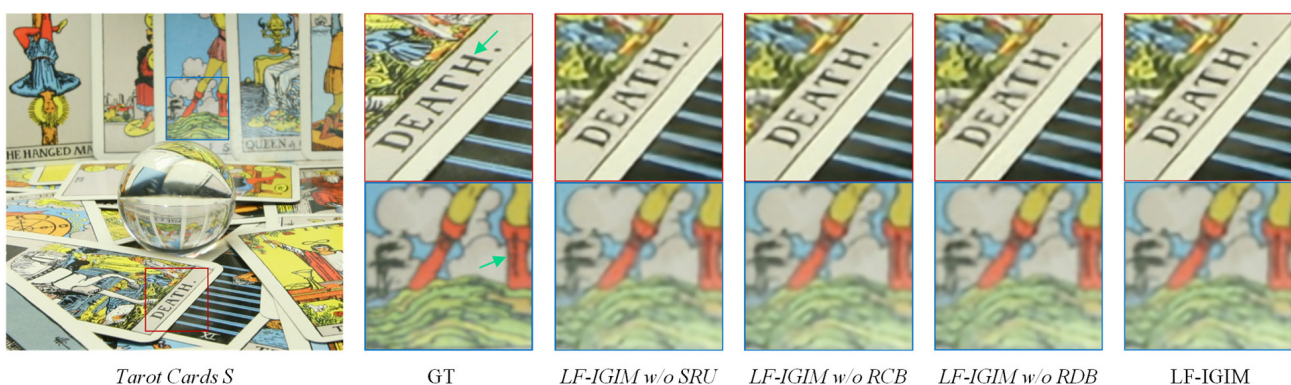


Figure 8. The 4×SR visualization results of the ablation study on *Tarot Card S*.

5. Conclusions

In this paper, we introduce a new method for LF SSR that combines view interaction and hierarchical feature fusion. The core idea of this method is to efficiently utilize the

complementary information between all SAIs by reducing the redundancy of LF features while preserving LF parallax structure by effectively fusing shallow and deep features. Specifically, the RCB and view interaction blocks proposed in this paper effectively solve the problem of the utilization of correlation between all views, and the hierarchical feature fusion solves the problem of LF parallax structure consistency. A comparison of this method with other existing SOTA methods on the same datasets shows that the LF-IGIM proposed in this paper achieves excellent performance in both quantitative and qualitative results under reasonable parameters and FLOPs. The modules in our LF-IGIM were rigorously validated using PSNR and SSIM obtained in ablation experiments. In addition, the quantitative results of experiments on both synthetic and real-world datasets indicate that our LF-IGIM has good robustness in all scenarios.

In the experimental results, it can be seen that our network needs to be improved in terms of the number of parameters. We also take into account the other limitations of the proposed LF-IGIM, which can only achieve significant results in terms of SSR. Therefore, on the one hand, we will explore modules that implement SSR with fewer parameters in the future. On the other hand, we will learn about Transformer and improve the network by incorporating a novel attention mechanism, so that the network can be effectively applied to SASR. This will allow it to be extended to more relevant tasks.

Author Contributions: Conceptualization, W.W. and Q.Z.; methodology, W.W. and C.M.; software, W.W.; validation, W.W., Q.Z., C.M., B.W. and Y.F.; formal analysis, T.Y.; investigation, Q.Z.; resources, Q.Z. and C.M.; data curation, B.W.; writing—original draft preparation, W.W.; writing—review and editing, W.W., Q.Z. and C.M.; visualization, W.W.; supervision, Q.Z. and B.W.; project administration, T.Y.; funding acquisition, Q.Z. and Y.F. All authors have read and agreed to the published version of the manuscript.

Funding: This work was funded by the National Key R&D Program of China under Grant 2023YFC3305802, the National Natural Science Foundation of China under Grant 62301320, the Natural Science Foundation of Fujian under Grant 2023J011009, and the Program for the Putian Science and Technology Bureau under Grant 2021G2001ptxy08.

Data Availability Statement: The experimental conditions and parameter settings of this study are described in detail in the summary of Section 4.1, but the code is related to subsequent research, so it will not be published for the time being. If you want to know the specific content of the code, you can contact us through email (1000467199@smail.shnu.edu.cn).

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Chen, Y.; Jiang, G.; Jiang, Z.; Yu, M.; Ho, Y.-S. Deep Light Field Super-Resolution Using Frequency Domain Analysis and Semantic Prior. *IEEE Trans. Multimed.* **2022**, *24*, 3722–3737. [[CrossRef](#)]
2. 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.* **2019**, *26*, 204–208. [[CrossRef](#)]
3. Zhang, S.; Shen, Z.; Lin, Y. Removing foreground occlusions in light field using micro-lens dynamic filter. In Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI), Montreal, QC, Canada, 19–27 August 2021; pp. 1302–1308.
4. Liu, N.; Zhao, W.; Zhang, D.; Han, J.; Shao, L. Light field saliency detection with dual local graph learning and reciprocative guidance. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada, 10–17 October 2021; pp. 4712–4721.
5. Chuchvara, A.; Barsi, A.; Gotchev, A. Fast and accurate depth estimation from sparse light fields. *IEEE Trans. Image Process.* **2020**, *29*, 2492–2506. [[CrossRef](#)]
6. Yu, J. A Light-Field Journey to Virtual Reality. *IEEE MultiMedia* **2017**, *24*, 104–112. [[CrossRef](#)]
7. 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.
8. Rossi, M.; Frossard, P. Geometry-Consistent Light Field Super-Resolution via Graph-Based Regularization. *IEEE Trans. Image Process.* **2018**, *27*, 4207–4218. [[CrossRef](#)]
9. Ghassab, V.K.; Bouguila, N. Light Field Super-Resolution Using Edge-Preserved Graph-Based Regularization. *IEEE Trans. Multimed.* **2020**, *22*, 1447–1457. [[CrossRef](#)]

10. Johannsen, O.; Honauer, K.; Goldluecke, B.; Alperovich, A.; Battisti, F.; Bok, Y.; Brizzi, M.; Carli, M.; Choe, G.; Diebold, M.; et al. A taxonomy and evaluation of dense light field depth estimation algorithms. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Honolulu, HI, USA, 21–26 July 2017; pp. 1795–1812.
11. 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 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020; pp. 2257–2266.
12. Zhang, S.; Lin, Y.; Sheng, H. Residual Networks for Light Field Image Super-Resolution. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019; pp. 11038–11047.
13. Zhou, L.; Gao, W.; Li, G.; Yuan, H.; Zhao, T.; Yue, G. Disentangled Feature Distillation for Light Field Super-Resolution with Degradations. In Proceedings of the 2023 IEEE International Conference on Multimedia and Expo Workshops (ICMEW), Brisbane, Australia, 10–14 July 2023; pp. 116–121.
14. Wang, Y.; Wang, L.; Yang, J.; An, W.; Yu, J.; Guo, Y. Spatial-Angular Interaction for Light Field Image Super-Resolution. In *Computer Vision—ECCV 2020: 16th European Conference, Glasgow, UK, 23–28 August 2020, Proceedings, Part XXIII 16*; Springer: Cham, Switzerland, 2020; pp. 290–308.
15. 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.* **2021**, *30*, 1057–1071. [[CrossRef](#)] [[PubMed](#)]
16. Cheng, Z.; Xiong, Z.; Liu, D. Light Field Super-Resolution by Jointly Exploiting Internal and External Similarities. *IEEE Trans. Circuits Syst. Video Technol.* **2020**, *30*, 2604–2616. [[CrossRef](#)]
17. Park, H.-J.; Shin, J.; Kim, H.; Koh, Y.J. Light Field Image Super-Resolution Based on Multilevel Structures. *IEEE Access* **2022**, *10*, 59135–59144. [[CrossRef](#)]
18. Wang, Y.; Wang, L.; Liang, Z.; Yang, J.; An, W.; Guo, Y. Occlusion-Aware Cost Constructor for Light Field Depth Estimation. In Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, 18–24 June 2022; pp. 19777–19786.
19. Wafa, A.; Pourazad, M.T.; Nasiopoulos, P. A Deep Learning Based Spatial Super-Resolution Approach for Light Field Content. *IEEE Access* **2021**, *9*, 2080–2092. [[CrossRef](#)]
20. 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](#)]
21. Wang, Y.; Lu, Y.; Wang, S.; Zhang, W.; Wang, Z. Local-Global Feature Aggregation for Light Field Image Super-Resolution. In Proceedings of the ICASSP 2022–2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Singapore, 22–27 May 2022; pp. 2160–2164.
22. Kar, A.; Nehra, S.; Mukhopadhyay, J.; Biswas, P.K. Sub-Aperture Feature Adaptation in Single Image Super-Resolution Model for Light Field Imaging. In Proceedings of the 2022 IEEE International Conference on Image Processing (ICIP), Bordeaux, France, 16–19 October 2022; pp. 3451–3455.
23. Chen, Y.; Jiang, G.; Yu, M.; Xu, H.; Ho, Y.-S. Deep Light Field Spatial Super-Resolution Using Heterogeneous Imaging. *IEEE Trans. Vis. Comput. Graph.* **2022**, *29*, 4183–4197. [[CrossRef](#)] [[PubMed](#)]
24. Cheng, Z.; Xiong, Z.; Chen, C.; Liu, D.; Zha, Z.-J. Light Field Super-Resolution with Zero-Shot Learning. In Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 20–25 June 2021; pp. 10005–10014.
25. Yao, H.; Ren, J.; Yan, X.; Ren, M. Cooperative Light-Field Image Super-Resolution Based on Multi-Modality Embedding and Fusion with Frequency Attention. *IEEE Signal Process. Lett.* **2022**, *29*, 548–552. [[CrossRef](#)]
26. Wang, S.; Zhou, T.; Lu, Y.; Di, H. Detail-Preserving Transformer for Light Field Image Super-Resolution. *AAAI Conf. Artif. Intell.* **2022**, *36*, 2522–2530. [[CrossRef](#)]
27. Yun, S.; Jang, J.; Paik, J. Geometry-Aware Light Field Angular Super Resolution Using Multiple Receptive Field Network. In Proceedings of the 2022 International Conference on Electronics, Information, and Communication (ICEIC), Jeju, Republic of Korea, 6–9 February 2022; pp. 1–3.
28. Liu, G.; Yue, H.; Wu, J.; Yang, J. Efficient Light Field Angular Super-Resolution with Sub-Aperture Feature Learning and Macro-Pixel Upsampling. *IEEE Trans. Multimed.* **2022**, *25*, 6588–6600. [[CrossRef](#)]
29. 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.* **2022**, *44*, 1819–1836. [[CrossRef](#)] [[PubMed](#)]
30. Zhou, L.; Gao, W.; Li, G. End-to-End Spatial-Angular Light Field Super-Resolution Using Parallax Structure Preservation Strategy. In Proceedings of the 2022 IEEE International Conference on Image Processing (ICIP), Bordeaux, France, 16–19 October 2022; pp. 3396–3400.
31. Duong, V.V.; Huu, T.N.; Yim, J.; Jeon, B. Light Field Image Super-Resolution Network via Joint Spatial-Angular and Epipolar Information. *IEEE Trans. Comput. Imaging* **2023**, *9*, 350–366. [[CrossRef](#)]
32. Sheng, H.; Wang, S.; Yang, D.; Cong, R.; Cui, Z.; Chen, R. Cross-View Recurrence-based Self-Supervised Super-Resolution of Light Field. *IEEE Trans. Circuits Syst. Video Technol.* **2023**, *33*, 7252–7266. [[CrossRef](#)]
33. Li, J.; Wen, Y.; He, L. SCConv: Spatial and Channel Reconstruction Convolution for Feature Redundancy. In Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Vancouver, BC, Canada, 17–24 June 2023; pp. 6153–6162.

34. Liu, G.; Yue, H.; Wu, J.; Yang, J. Intra-Inter View Interaction Network for Light Field Image Super-Resolution. *IEEE Trans. Multimed.* **2023**, *25*, 256–266. [[CrossRef](#)]
35. Rerábek, M.; Ebrahimi, T. New light field image dataset. In Proceedings of the International Conference on Quality of Multimedia Experience, Lisbon, Portugal, 6–8 June 2016; pp. 1–2.
36. Honauer, K.; Johannsen, O.; Kondermann, D.; Goldluecke, B. A Dataset and Evaluation Methodology for Depth Estimation on 4D Light Field. In *Computer Vision—ACCV 2016: 13th Asian Conference on Computer Vision, Taipei, Taiwan, 20–24 November 2016, Revised Selected Papers, Part III 13*; Springer: Cham, Switzerland, 2016; pp. 19–34.
37. Wanner, S.; Meister, S.; Guillemot, C. Datasets and Benchmarks for Densely Sampled 4D Light Fields. *Vis. Model. Vis.* **2013**, *13*, 225–226.
38. Pendu, M.L.; Jiang, X.; Guillemot, C. Light Field Inpainting Propagation via Low Rank Matrix Completion. *IEEE Trans. Image Process.* **2018**, *27*, 1981–1993. [[CrossRef](#)] [[PubMed](#)]
39. Vaish, V.; Adams, A. *The (New) Stanford Light Field Archive*; Computer Graphics Laboratory, Stanford University: Stanford, CA, USA, 2008; Volume 6.

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.