Article

Light Field Image Quality Enhancement by a Lightweight Deformable Deep Learning Framework for Intelligent Transportation Systems

David Augusto Ribeiro [1], Juan Casavílca Silva [2], Renata Lopes Rosa [1], Muhammad Saadi [3], Shahid Mumtaz [4], Lunchakorn Wuttisittikulkij [5,*], Demóstenes Zegarra Rodríguez [1] and Sattam Al Otaibi [6]

1. Department of Computer Science, Federal University of Lavras, Minas Gerais 37200-000, Brazil; ribeiro@estudante.ufla.br (D.A.R.); renata.rosa@ufla.br (R.L.R.); demostenes.zegarra@ufla.br (D.Z.R.)
2. Department of Sciences, Pontifical Catholic University of Peru, Lima 1801, Peru; casavilca.je@pucp.edu.pe
3. Department of Electrical Engineering, Faculty of Engineering, University of Central Punjab, Lahore 54000, Pakistan; muhammad.saadi@ucp.edu.pk
4. Instituto de Telecomunicaces, 3750011 Aveiro, Portugal; smumtaz@av.it.pt
5. Wireless Communication Ecosystem Research Unit, Department of Electrical Engineering, Chulalongkorn University, Bangkok 10330, Thailand
6. Department of Electrical Engineering, College of Engineering, Taif University, Taif 21944, Saudi Arabia; srotaibi@tu.edu.sa
* Correspondence: lunchakorn.w@chula.ac.th; Tel.: +66-866-889-858

Abstract: Light field (LF) imaging has multi-view properties that help to create many applications that include auto-refocusing, depth estimation and 3D reconstruction of images, which are required particularly for intelligent transportation systems (ITSs). However, cameras can present a limited angular resolution, becoming a bottleneck in vision applications. Thus, there is a challenge to incorporate angular data due to disparities in the LF images. In recent years, different machine learning algorithms have been applied to both image processing and ITS research areas for different purposes. In this work, a Lightweight Deformable Deep Learning Framework is implemented, in which the problem of disparity into LF images is treated. To this end, an angular alignment module and a soft activation function into the Convolutional Neural Network (CNN) are implemented. For performance assessment, the proposed solution is compared with recent state-of-the-art methods using different LF datasets, each one with specific characteristics. Experimental results demonstrated that the proposed solution achieved a better performance than the other methods. The image quality results obtained outperform state-of-the-art LF image reconstruction methods. Furthermore, our model presents a lower computational complexity, decreasing the execution time.

Keywords: light field imaging; deep learning framework; image quality; computational complexity; intelligent transportation systems

1. Introduction

A light field describes the distribution of light rays in the space; thus, more information from our environment can be used to build an image. However, due to the high dimensionality of the data, to obtain a scene is a difficult task [1].

Currently, the Light Field (LF) imaging [1] area has been explored by many studies [2,3] in the field of Virtual Reality (VR), Augmented Reality (AR) and different industrial applications, such as the commercial plenoptic cameras. In addition, different image-based solutions are used in Intelligent Transportation Systems (ITSs) for several applications [4–10], which use different machine learning techniques. ITS solutions aim to improve safety, mobility and efficiency of transport services, and to accomplish these goals, visual information plays an important role in the development of these services. Nowadays, there are many proposals of deep learning models [11–19] that are applied to

image or video processing applications in order to obtain better performance results in terms of classification accuracy or used to improve the image content quality.

Research on microlens array (MLA) [20] has involved the intertwining with the LF imaging with the aim to obtain four-dimensional information. Thus, with the microlens array between the main lens and the image sensor, the plenoptic cameras can capture the direction information of rays and intensity in real-world scenes [21]. However, some problems with the trade-off between angular and spatial resolution cause conflicts in dividing limited sensor resolution and restrictions on LF imaging. In some image scenes, the use of more information of spatial resolution in relation to angular resolution may present some advantages, but in other scenes more angular resolution may be more useful. Spatial resolution is an important factor for visual realism in 2D and 3D displays [22]. Angular resolution is related to the quality of the parallax effect, if it is insufficient, important details of the presented information negatively impact the viewing experience. Thus, both spatial and angular resolutions contain relevant properties to obtain a robust LF image content.

For improving the angular resolution of LF images, the view synthesis can be performed, and Sub-Aperture Images (SAIs) can be synthesized from a set of input views extracted from the plenoptic cameras. Some view synthesis studies [23,24] break down the view synthesis into the disparity estimator and a color predictor is performed by a Convolutional Neural Network (CNN) [25–28]. These studies obtain better results than other methods [29,30] that require depth information for view warping and image registration. However, there are limitations in [23,24] for reconstructing LF scenes [31] as the non-Lambertian surfaces and occluded regions.

The view methods are commonly concerned with the geometry data of the scenes, with inaccurate depth estimation [32]. In [33], the advantage of the clear texture structure of the echo planer image (EPI) is taken in the LF data, and the study models the problem of LF reconstruction from a sparse set of views as a CNN-based angular detail restoration on EPI. A blur–restoration–deblur framework is proposed, and the estimation of the geometric scene is not performed, and the EPIs are 2D slices of 4D LF. Additionally, in [33], the proposal is highly time-consuming because the task of blur–restoration–deblur is executed several times before synthesization. In [34], an extension of the framework of [33] is performed, in which ghosting effects are suppressed. However, the structure of both framework are equal.

Despite the large advances in the image processing research area in the last few years, many challenges need to be addressed [35–40]. The LF imaging requires the ability to capture higher dimensional data as opposed to simply recording a 2D projection in photography. Additionally, for acquiring a high-dimensional data and resolution, a trade-off between the dimensions is imposed; therefore, the computational complexity is also a challenge for the design of the algorithms.

In addition, 4D LF data are correlated in ray space, containing abundant information on the scene and consequent computational complex. Unlike the manipulation of 2D matrix or 3D volume-like videos, the high-dimensional LF data manipulation with plain CNN becomes a difficult challenge because of the amount of data to be processed.

Many times, high-resolution (HR) images are required in various LF applications [41]. Thus, it is necessary to perform the reconstruction of HR images from low-resolution (LR) data to LF super-resolution (SR) image [42]. For obtaining an acceptable SR performance, information containing a single view, such as spatial data and different views, such as angular information is very important. A few models have been proposed in this area [43–46]; however, these studies are limited due to their poor spatial information data.

Other methods using deep-learning-based algorithms [47–51] have improved the spatial information data through convolutions layers, obtaining better results as compared to traditional methods. However, these studies have not resolved the problem of disparity in LF image SR.

In this context, the present research proposes a lightweight deformable convolution network to treat the problem of disparity in LF SR images. An angular alignment approach

is performed for angular data incorporation. Feature extraction containing rich spatial data is performed to align with their original features, and a soft activation function is also used in the CNN model to decrease the computational complexity of the proposed deformable deep learning framework. Consequently, the proposed framework obtains an improvement on the final image quality.

The main contributions of this paper are listed below.

1. An improved framework, which considers the feature extraction and angular alignment using the deformable convolution network approach, ruling out the use of applying a loss function.
2. To reduce the computational complexity for LF SR images, a novel activation function is utilized, which is performed in the proposed CNN model. Thus, a lightweight solution to process LF SR images is obtained.
3. The performance assessment of the proposed model is tested using recent databases. Experimental results demonstrated that our proposal reached a high accuracy for image reconstruction, obtaining a better performance in image quality than other similar works.
4. Our proposed framework improves the image content and its perceptual quality, which are obtained with a reduced computational processing and execution time that is relevant for different applications in the ITS research area [52–54].

Experimental results showed that our proposed CNN architecture obtain a low computational complexity, reducing, on average, 37% of the training time and, on average, 40% of the execution time. Moreover, image quality was also evaluated, and the results demonstrated a superior performance of the proposed model in terms of objective image quality metrics, such as Structural Similarity Index Measure (SSIM) and peak signal-to-noise ratio (PSNR), reaching score values of 0.99 and superior to 45, respectively.

The remainder of this paper is organized as follows. In Section 2, related works are presented. The methodology and the details of the proposed method is presented in Section 3. Experimental results are presented in Section 4. Finally, the conclusions are presented in Section 5.

2. Related Works

In this section, some works about LF image representation, as welll as frameworks based on Deep Learning algorithms, are treated.

2.1. Light Field Representation and Images

The most common solution for the representation of a 4D LF is the light rays parameterized by the coordinates of their intersections with two planes in arbitrary positions. Thus, the coordinate system is represented by (u, v) for the first plane, and (s, t) is the representation for the second one.

The plenoptic function that describes a LF is reduced from seven to only four dimensions, and it is represented by Equation (1).

$$L(u, v, s, t) \qquad (1)$$

A 4D LF can be visualized in two ways, through an integral LF structure, and 2D slices [55,56]. Thus, the 4D LF can be represented as being a 2D array of images. For LF rendering, the capture of insufficient samples can cause the ghosting effect in the views. However, it is impractical to acquire many samples of a LF [1]. The minimum number of samples needed for light field rendering is studied in [57,58], which concluded that the pixels must at least touch each other to render the views without producing the ghosting effect. Thus, a large number of samples are needed for producing a noise-free output, what is computationally expensive, even now.

Many methods and models have been developed for working with LF images. An approach was developed in [59], to estimate disparity from a LF images. Farrugia et al. [43]

proposed a linear subspace projection approach for LF image SR. In [60], a LFBM5D for LF image denoising is proposed, extending the state-of-the-art Block-matching and 3D filtering (BM3D) image denoising filter to LFs. Another method was used to achieve LF image SR in [44], using a graph-based method via graph optimization. Although the LF images are well encoded in these cited studies, the spatial information is not fully exploited. Recently, deep learning methods [61] are achieving superior results when compared to traditional methods in spatial information exploitation. However, the computational models are much more complex and time-consuming for processing.

In our work, the feature extraction and angular alignment are performed to improve the image quality, reducing noise effects, and a soft activation function was used in the CNN model for decreasing computational expenses.

2.2. Frameworks Using Deep Learning Algorithms

Deep-learning methods [62–64] have been used for several applications [65,66], such as classification, detection, and recognition of images. For Single Image Super-Resolution (SISR), a framework is proposed in [67], which learns the mapping from LR to HR image using three layers, patch representation, non-linear mapping and reconstruction. Dong et al. [68] use the SRCNN structure [67] to achieve a speed up of more than 40 times with even superior restoration quality. A DRCN structure is proposed in [69], which improves the SR results without introducing new parameters. Lai et al. [70] propose a method that adopts a Laplacian pyramid to reconstruct residuals of high-resolution images. Hu et al. [71] propose a method to solve SISR of arbitrary scale factor with a single model. The cited studies work on obtaining a high-resolution image. However, they still have a large computational expense.

Currently, novel SISR methods are demonstrating superior performance to traditional methods in spatial information exploitation. The LFCNN approach is used in [72], improving both the efficiency of training and the quality of angular SR results by using weight sharing. In [73], the authors attempt to measure the degree of their LF coherence (LFC), obtaining consistent performance. Yuan et al. [47] use the LF-DCNN model for improving the LFCNN via a SISR network EDSR [74] and a specific EPI-enhancement network.

A bidirectional recurrent network LFNet is proposed in [49] by extending BRCN to LFs. Wang et al. [75] proposed another method, named LF-InterNet, for interacting spatial and angular information for LF image SR. LF-ATO [76] and LF-InterNet [75] has achieved a high reconstruction accuracy. Although the recent studies have improved the network performance, the problem of disparity problem has not been well explored in the literature. In the LFSSR [50] and LF-InterNet model [75], the LF features are organized, and the angular information is incorporated in the model. However, the disparity problem continues to occur in these studies. The LFNet [49] works with a video SR framework to address the problem of disparity in recurrent networks, but it considers only SAIs from the same row or column as its inputs.

The configuration in regular CNNs, which consider a fixed kernel, does not explore long-range information. For resolving this problem, a deformable convolution is proposed in [77] considering additional and learned offsets to make the convolution kernel distant from its neighborhood. However, the deformable convolutions have been applied to video SR [78,79] or more complex computational systems [77].

3. Methodology

In this section, the main steps followed in building the proposed framework are described. We introduce the framework topology, used datasets and evaluation of the proposed method through comparison to other methods.

3.1. Proposed Framework

Figure 1 shows the topology of the Lightweight Deformable Deep Learning Framework, including the feature extraction, angular alignment (AA) using the deformable convolution approach, and the reconstruction step. The input LR data serve as input in

the CNN model, which performs the feature extraction and, posteriorly, the AA using the deformable approach; then, reconstruction is performed. The deformable convolution network approach considers constrained pooling layer models to treat the information related to angular resolution in order to improve the image content and perceptual quality. By the end, the reconstructed data are generated, in which the LF data are represented as $L(x, y, s, t)$.
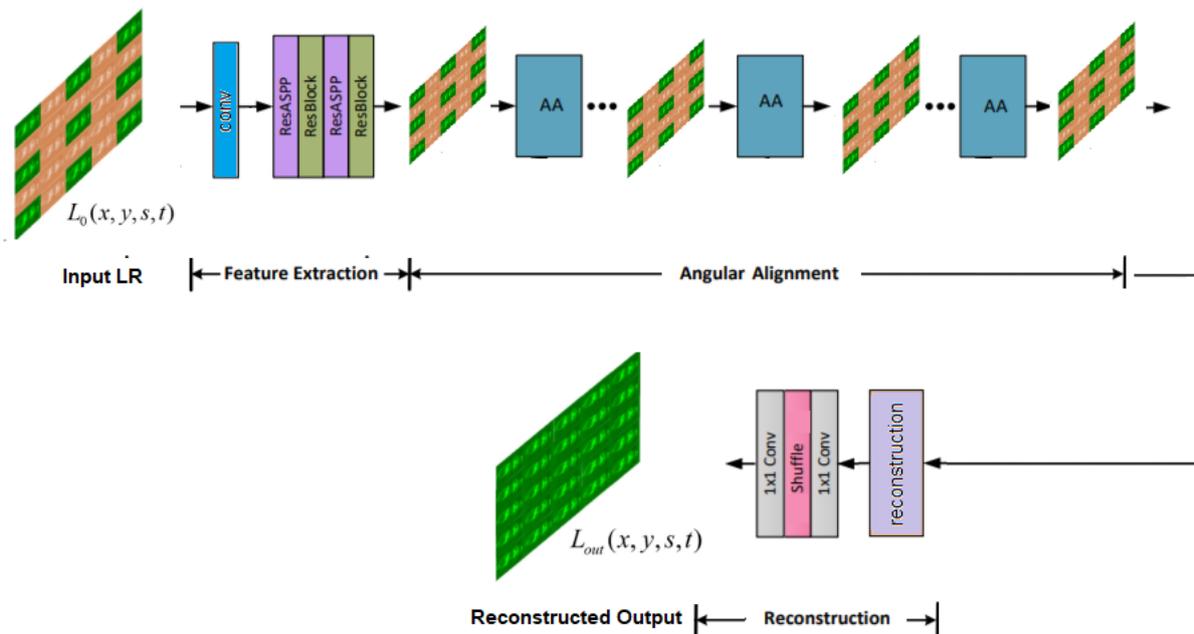

Figure 1. An overview of the proposed lightweight deformable deep learning framework.

3.1.1. Feature Extraction

The feature representation containing a rich spatial context information is useful to the subsequent alignments and reconstruction steps. Thus, in this work, spatial pyramid pooling is used for performing the feature extraction.

The inputs are processed with a 1×1 convolution, for generating initial features. The residual modules and blocks are used for performing deep feature extraction. Then, 3×3 convolutions are combined in the residual blocks. Later, features of these branches are added in 1×1 convolution.

The activation function used in this work is defined by Equation (2).

$$SR(t) = \frac{t}{\frac{t}{\alpha} + e^{-\frac{t}{\beta}}} \tag{2}$$

in which, α and β are a pair of trainable positive parameters. The activation function presents a non-monotonic region, and $t < 0$ has the property with zero mean. In the case of $t > 0$, it avoids and rectifies the output distribution.

In the experiments, other activation functions are used, such as Leaky ReLU for comparison with the SR function.

3.1.2. Angular Alignment

After the feature extraction, an angular alignment using a deformable convolution network approach is performed, in which a bidirectional alignment incorporates angular data. Side-view features are put to the center view, and then they are aligned with the center-view feature. In this work, a deformable convolution occurs for performing the feature

collection and another for distribution. The first convolution considers the $(k-1)$th side-view $\mathcal{R}_i^{k-1}$ and offsets ΔP_i^k for generating the k-th feature $\mathcal{R}_{i \to c}^k$, as shown in Equation (3).

$$\mathcal{R}_{i \to c}^k = H_{dcn}^k \left(\mathcal{R}_i^{k-1}, \Delta P_i^k \right), \tag{3}$$

where H_{dcn}^k is the deformable convolution in the kth block, $\Delta P_i^k = \{\Delta p_n\} \in \mathbb{R}^{H \times W \times C'}$ represents the offset of $\mathcal{R}_i^{k-1}$ from $\mathcal{R}_c$.

An offset generation branch is used in this work, learning the offset ΔP_i^k. The side-view feature $\mathcal{R}_i^{k-1}$ is added to the center-view feature $\mathcal{R}_c$, going to a 1×1 convolution for performing a feature reduction. After, a residual module is applied to enlarge the receptive field, maintaining a dense sampling rate. Thus, the residual module improves the angular dependencies between the center and side views. By the end, another 1×1 convolution is used for generating an offset feature.

A 1×1 convolution is performed, adding the angular data in the aligned features.

$$\mathcal{R}_c^k = H_{1 \times 1}^k \left(\left[\mathcal{R}_{1 \to c}^k, \mathcal{R}_{2 \to c}^k, \cdots, \mathcal{R}_{(A^2-1) \to c}^k, \mathcal{R}_c \right] \right), \tag{4}$$

where $[\cdot, \cdot]$ represents the concatenation, and $H_{1 \times 1}^k$ represents the 1×1 convolution.

To super-resolve all LF images, the incorporated angular information need to be encoded into each side view. Consequently, we perform feature distribution to propagate the incorporated angular information to the side views. Since the disparities between the side-view features and center-view features are mutual, we do not perform additional offset learning. Instead, we use the opposite offset $\Delta \bar{P}_i^k = -\Delta P_i^k$ to warp the fused center-view feature $\mathcal{R}_c^k$ to the i-th side view. That is,

$$\mathcal{R}_i^k = H_{dcn}^k \left(\mathcal{R}_c^k, \Delta \bar{P}_i^k \right). \tag{5}$$

Posteriorly, the center-view feature $\mathcal{R}_c^k$ and side-view $\mathcal{R}_i^k, (i = 1, 2, \cdots, A^2 - 1)$ are generated by the k-th.

In the proposed model, the alignment is performed among the center views and each side view. It is important to note that the number of alignments can influence the network model. Thus, the performance of the proposed model was analyzed according to the variations' number of alignments.

3.1.3. Reconstruction

For high reconstruction accuracy, spatial and angular data are used in the framework, and a reconstruction step was necessary to add the features for the LF image. Thus, multi-distillation blocks are used with a mechanism to extract and process hierarchical features with the aim to achieve a small number of parameters and, consequently, a low computational cost.

The outputs of the feature extraction and each alignment are processed by a 1×1 convolution. The coarsely fused feature goes to the stacked information blocks. In each information block, the input feature is processed by a 3×3 convolution and an activation function.

The narrow feature fed to the bottleneck of the information block and the wide feature goes to a 3×3 convolution. Posteriorly, features of different stages are processed by a 1×1 convolution, and the feature of the last information block is processed by a 3×3 convolution for reducing its depth from 128 to 32.

A 1×1 convolution is used for the reconstructed features, extending the depth to $\alpha^2 C$. The α is an upsampling factor. A pixel shuffle is used for upscaling the reconstructed feature, with a resolution $\alpha H \times \alpha W$. Thus, a 1×1 convolution is used to compress the number of feature channels.

Moreover, we have justified in additional experiments that the detail–restoration network can be certainly substituted by a deeper or more complex network structures, which will further improve the performance of LF reconstruction.

3.2. Model of the Network

In this work, input sparse views $S_0(x, y, s, t)$ with the resolution of (H, W, n, n) are used and one angular dimension $t = t^*, t^* \in \{1, 2, ..., n\}$ extracts 3D volume, containing a resolution of (H, W, n) as shown in Equation (6).

$$Bl_{t*}(x, y, s) = L_0(x, y, s, t^*) \tag{6}$$

$Bl_{t*}(x, y, s)$ are interspersed as $Bl_{t*}(x, y, s) \uparrow$ to the resolution (H, W, N). Thus, the details of $Bl_{t*}(x, y, s) \uparrow$ are restored as $F_{r3d}(\cdot)$, forming the intermediate LF in Equation (7).

$$Bl_{inter}(x, y, s, t^*) = F_{r3d}(Bl_{t*}(x, y, s) \uparrow) \tag{7}$$

An angular domain conversion is performed to transform from t to dimension s. Using $s = s^*, s^* \in \{1, 2, ..., N\}$ are extracted from $Sl_{inter}(x, y, s^*, t)$ as is shown in Equation (8).

$$Bl_{s*}(x, y, t) = Sl_{inter}(x, y, s^*, t) \tag{8}$$

The resolution of (H, W, n) is interspersed to $Bl_{s*}(x, y, t) \uparrow$ at same resolution in $Bl_{t*}(x, y, s) \uparrow$. Thus, the detail–restoration network is used for recovering details of $Bl_{s*}(x, y, t) \uparrow$, as $F_{c3d}(\cdot)$. The output $Sl_{out}(x, y, s, t)$ as the resolution of (H, W, N, N) is shown in Equation (9).

$$Sl_{out}(x, y, s^*, t) = Fl_{c3d}(Bl_{s*}(x, y, t) \uparrow) \tag{9}$$

3.3. Details of Implementation of the CNN Model

In this work, a model of an angular resolution of 5×5 was used. The learning rate of our model was set to 4×10^{-4}, and then it was decreased by a factor of 0.5. This occurred for every 10 epochs. The training phase finished at 50 epochs.

The optimization of the training of the CNN model is performed by the mini-batch momentum Stochastic Gradient Descent (SGD) approach, and the filters of the CNN are initialized through a zero-mean Gaussian distribution.

Our model was implemented using the deep learning API written in Python called Keras, on a workstation with an Intel 3.6 GHz CPU and a TiTan X GPU.

Tests are performed with the SR function and, for comparison, we also used the well-known activation function, Leaky ReLU.

3.4. Datasets

Some public LF datasets, such as INRIA [80], HCInew [81], EPFL [82], and HCIold [83], were used in this work. They are presented in Table 1 with the main characteristics. These datasets were chosen because they are the most used in the related works [24,75].

Table 1 presents the number of scenes for training and for testing of each dataset used in this work. Each dataset presents a total number of scenes available, represented by the column-named Scenes. The LFs of the datasets have an angular resolution, AngRes, of 9×9. In the training stage, each SAI was cropped into HR patches with stride of 32. The bicubic downsampling approach was used for generating the LR patches containing a resolution of 64×64. It is important to note that a random horizontal and vertical flipping, 90-degree rotation was performed in this work, augmenting the training data by eight times.

Table 1. Public datasets used in this work and its characteristics.

Dataset	Training	Test	Type	Scenes	AngRes	SpaRes (Mpx)	GT Depth
EPFL[a] [82]	70	10	real (lytro)	119	14×14	0.034	no
HCInew [b] [81]	20	4	synthetic	24	9×9	0.026	yes
HCIold [c] [83]	10	2	synthetic	12	9×9	0.070	yes
INRIA [d] [80]	35	5	real (lytro)	57	14×14	0.027	no

[a] École polytechnique fédérale de Lausanne, [b] Heidelberg Collaboratory for image processing—new DB, [c] Heidelberg Collaboratory for image processing—old DB, [d] French National Institute for Research in Computer Science and Control.

3.5. Evaluation of the Proposed Method through Comparison with Others' Methods

Our method was compared to some state-of-the-art methods, including single image methods, such as EDSR [74], RCAN [84], and SAN [85], and some LF image SR methods, such as LFNet [49], LFSSR [50], resLF [48], LF-ATO [76], and LF-InterNet [75].

According to the related works [48,49,75], the Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM) were used as quantitative metrics for image quality assessment.

The PSNR is determined using the following relation:

$$PSNR = 10 \log_{10} \frac{(2^d - 1)^2 WH}{\sum_{i=1}^{W} \sum_{j=1}^{H} (p[i,j] - p'[i,j])^2} \quad (10)$$

where d represents the bit depth of pixel, W represents the image width, H is the image height, and $p[i,j]$, $p'[i,j]$ represent the ith-row jth-column pixel in the original and reconstructed image, respectively.

The SSIM is computed by:

$$SSIM(P) = \frac{2*\mu_1(P)*\mu_2(P)+C1}{\mu_1(P)^2+\mu_2(P)^2+C1} * \frac{2*cov(P)+C2}{s1(P)^2+s2(P)^2+C2} \quad (11)$$

where $\mu_1(P)$ and $\mu_2(P)$ represent the mean value of seq1 and seq2 computed in a window located around P image; $s1(P)$ and $s2(P)$ represent the standard deviation of $seq1$ and $seq2$ computed over the same window; $cov(P)$ is the covariance between $seq1$ and $seq2$ computed over the same window; $C1 = (K1 * L)^2$ and $C2 = (K2 * L)^2$ represent the regularization constants, in which $K1$, $K2$ are the regularization parameters, and they must be > 0; L is the dynamic range of the pixel values.

For measuring the computational efficiency, our proposed method was compared to same methods used in the image quality assessment. For this performance comparison, the number of parameters, #Params, for measuring the model size, and the FLOPs for measuring the memory cost were captured.

Additionally, the training and execution time of our proposed method and other related proposals are measured. It is important to note that the efficiency of our proposed method will be measured for $4 \times$ SR scale.

4. Experimental Results

In this section, the main results about the use of AA in the network model, and the proposed model performance compared to other state-of-the-art models are presented.

4.1. Angular Alignment in the Network Model

In this subsection, we investigate the tests for definition of the network model of our proposed solution, through the AA.

The relation between the number of alignments (#AA), average PSNR (Avg PSNR), and average SSIM (Avg SSIM) are studied; this is shown in Figure 2. Here, the average values of each dataset: INRIA [80], HCInew [81], EPFL [82], and HCIold [83] are shown.
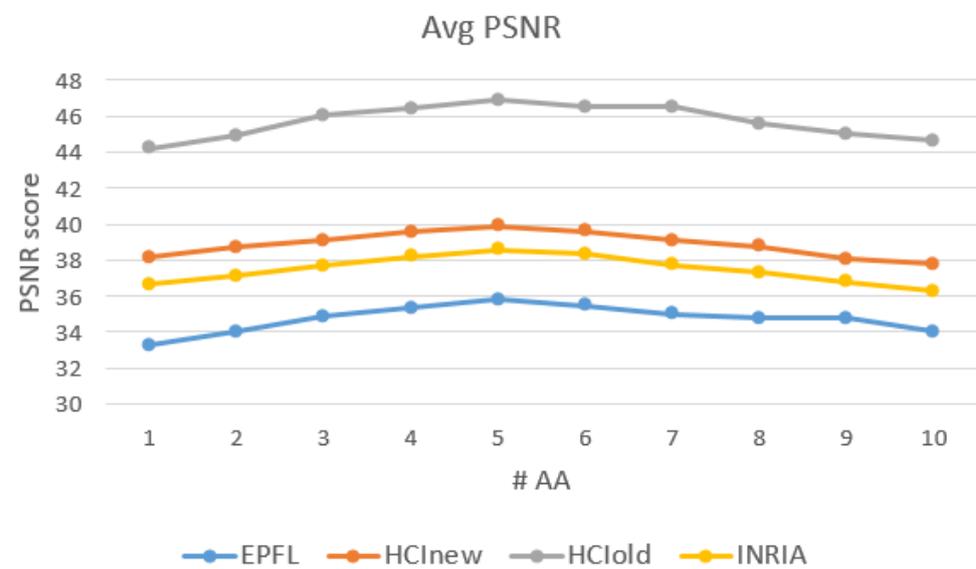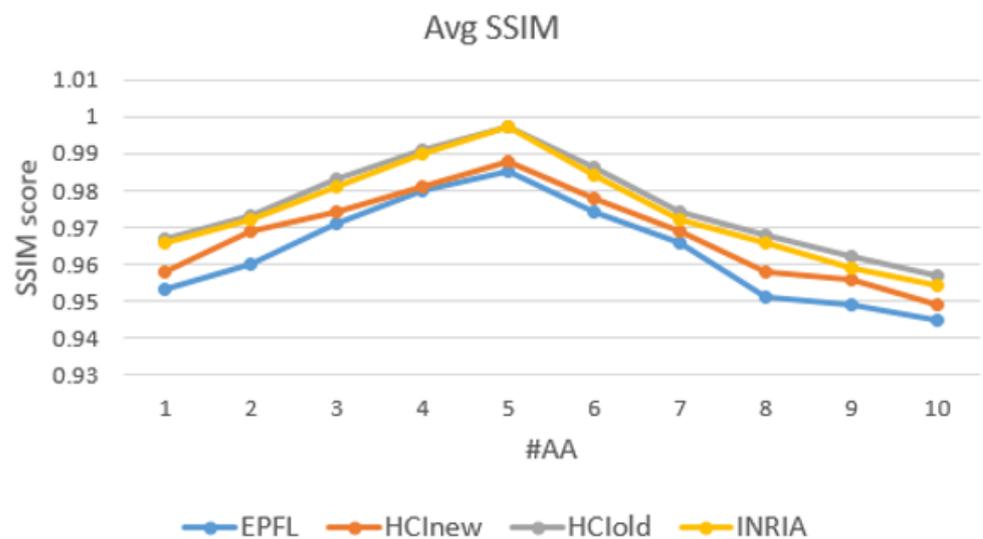
Figure 2. Relation between the number of alignments and average image quality scores using PSNR and SSIM, which were applied to the datasets INRIA [80], HCInew [81], EPFL [82], and HCIold [83]. (**a**) PSNR, and (**b**) SSIM.

It can be observed from Figure 2a,b that the number of AA converged to the value of five for both metrics PSNR and SSIM. It is important to note that the number of alignments represents the deformable convolutions in the feature distribution step, being an important role a scenario of LF image SR. The reconstruction accuracy is improved in the moment that the number of AA increases. However, the performance saturated in the #AA = 5.

4.2. Image Quality Assessment

Images generated by our proposed method and related methods are shown in Figure 3, which were generated using an image extracted from the INRIA [80] dataset.
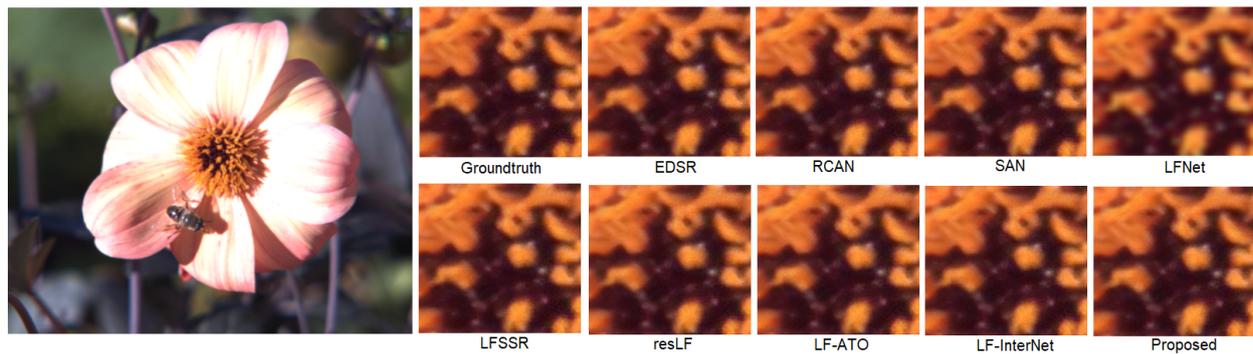
Figure 3. Visual results of the methods and the proposed one of $4 \times$ SR.

It is important to note that the perceptual quality of the images generated by our proposed model, considering $4 \times$ SR scale is compared with the groundtruth images.

The image quality assessment is quantitatively evaluated using objective metrics. In this test, all the images available in each dataset were used. Table 2 presents the PSNR scores obtained by our proposed model and other methods used for performance comparison.

Table 2. PSNR values achieved by different methods and our proposed model model in different datasets.

Method	EPFL	HCInew	HCIold	INRIA
EDSR [74]	33.01	35.29	42.01	34.33
RCAN [84]	34.22	35.02	42.14	35.12
SAN [85]	33.11	35.39	42.41	34.43
LFNet [49]	32.09	34.01	40.17	33.02
LFSSR [50]	35.19	37.23	44.11	37.37
resLF [48]	33.49	36.11	43.19	34.33
LF-ATO [76]	34.10	38.03	44.29	36.21
LF-InterNet [75]	34.36	38.09	45.33	36.37
Proposed model with Leaky ReLU	34.41	38.22	45.49	37.51
Proposed model with SR	35.83	39.91	46.89	38.59

As can be observed in Table 2, the proposed model with the SR activation function achieved the highest PSNR scores.

Similar results are obtained using SSIM, in which our proposed method achieved the best performance as can be observed in Table 3.

Table 3. SSIM values achieved by different methods and our proposed model in different datasets.

Method	EPFL	HCInew	HCIold	INRIA
EDSR [74]	0.943	0.940	0.960	0.942
RCAN [84]	0.945	0.942	0.962	0.948
SAN [85]	0.947	0.942	0.963	0.948
LFNet [49]	0.940	0.936	0.964	0.940
LFSSR [50]	0.951	0.949	0.963	0.951
resLF [48]	0.943	0.944	0.960	0.952
LF-ATO [76]	0.950	0.952	0.961	0.964
LF-InterNet [75]	0.950	0.950	0.964	0.964
Proposed model with Leaky ReLU	0.953	0.956	0.964	0.966
Proposed model with SR	0.985	0.988	0.997	0.997

4.3. Computational Efficiency

The comparison of our method to other methods was performed in terms of the number of parameters, #Params, and FLOPs in GFLOPs unit, whose results are shown in Table 4. As can be seen, our method uses a small number of parameters and a medium number of FLOPs.

Table 4. Comparisons of our method to others, using number of parameters, #Params, as well as FLOPs, using the SR activation function.

Method	#Params.	FLOPs (G)
EDSR [74]	14.18 M	15.33×25
RCAN [84]	14.39 M	15.71×25
SAN [85]	14.56 M	16.05×25
LFNet [49]	5.83 M	36.18
LFSSR [50]	6.23 M	36.87
resLF [48]	6.29 M	36.96
LF-ATO [76]	1.39 M	569.33
LF-InterNet [75]	4.58 M	46.18
Proposed model	3.17 M	43.41

In addition, the simulation time of our proposed model is compared to other methods, and the results show that the training takes to converge, using the activation function Leaky ReLU, around 7 hours and, using our method with the SR activation function, takes around 5 h. Thus, Table 5 shows a reduction, on average, of 37% of the training and, on average, of 40% of the execution time using the SR activation function when compared to the related works. The execution time is measured as an average for running in the datasets LF datasets, such as INRIA [80], HCInew [81], EPFL [82], and HCIold [83]. It is worth noting that all methods used for performance comparison were run on the same workstation with an Intel 3.6 GHz CPU and a TiTan X GPU.

Table 5. Simulation times achieved by different methods and the proposed one for training and to execute in the LF datasets, such as INRIA [80], HCInew [81], EPFL [82], and HCIold [83].

Method	Training (h)	Execution (h)
EDSR [74]	8.2	0.9
RCAN [84]	8.3	0.9
SAN [85]	9.1	1.1
LFNet [49]	9.8	1.3
LFSSR [50]	9.7	1.3
resLF [48]	8.9	1.1
LF-ATO [76]	8.4	1.0
LF-InterNet [75]	8.3	0.9
Proposed model with Leaky ReLU	7.2	0.7
Proposed model with SR	5.1	0.6

5. Conclusions

In this work, we propose a new method for better visual quality, and to decrease the problem of disparity in LF images. The procedure of feature alignment incorporates angular data as well as improves the image quality. Moreover, the experimental results verified the benefits of the proposed framework for the problem of depth estimation into LF images. In order to obtain more reliable and representative results, all methods used for comparison purposes were evaluated in datasets with different characteristics. Experimental results showed that our proposed framework obtained the best performance in relation to other methods. This fact demonstrated the versatility and good response in different image conditions. For the training and execution time of our proposed model, we verified a

reduction, on average, of 37% of the training and, on average, of 40% of the execution time using the SR activation function when compared to the related works. For the PSNR metric, we achieved values of 46.89 and, for the SSIM metric, we achieved values of 0.997 in the determined dataset used in this work. Such values were achieved by the learning capacity through the multi-scale feature representation and the activation function application.

The choice of the number of AA proved to have great importance in this work. Additionally, to increase the performance in relation to the training and execution time, the training function, SR, played a very important role in the network model.

In future work, we intend to explore the proposed model in the area of Biometrics LF Data, comparing the efficiency of our proposed model and recent state-of-the-art approaches. In addition, different activation functions will be tested to continue decreasing the time during the training and execution phases of the proposed framework.

Author Contributions: Conceptualization, D.A.R.; Formal analysis, D.A.R. and R.L.R.; Funding acquisition, S.A.O.; Investigation, J.C.S.; Methodology, J.C.S. and L.W.; Project administration, D.Z.R. and S.A.O.; Validation, M.S.; Visualization, S.M.; Writing—original draft, R.L.R. and D.Z.R.; Writing—review and editing, M.S., S.M. and D.Z.R. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest: The authors declare that they have no conflict of interest.

References

1. Wu, G.; Masia, B.; Jarabo, A.; Zhang, Y.; Wang, L.; Dai, Q.; Chai, T.; Liu, Y. Light Field Image Processing: An Overview. *IEEE J. Sel. Top. Signal Process.* **2017**, *11*, 926–954. [CrossRef]
2. Park, M.K.; Park, C.S.; Hwang, Y.S.; Kim, E.S.; Choi, D.Y.; Lee, S.S. Virtual-Moving Metalens Array Enabling Light-Field Imaging with Enhanced Resolution. *Adv. Opt. Mater.* **2020**, *8*, 2000820. [CrossRef]
3. Zhou, W.; Liu, G.; Shi, J.; Zhang, H.; Dai, G. Depth-guided view synthesis for light field reconstruction from a single image. *Image Vis. Comput.* **2020**, *95*, 103874. [CrossRef]
4. Haydari, A.; Yilmaz, Y. Deep Reinforcement Learning for Intelligent Transportation Systems: A Survey. *IEEE Trans. Intell. Transp. Syst.* **2020**, 1–22. [CrossRef]
5. Liang, X.; Zhang, Y.; Wang, G.; Xu, S. A Deep Learning Model for Transportation Mode Detection Based on Smartphone Sensing Data. *IEEE Trans. Intell. Transp. Syst.* **2020**, *21*, 5223–5235. [CrossRef]
6. Kumar, N.; Rahman, S.S.; Dhakad, N. Fuzzy Inference Enabled Deep Reinforcement Learning-Based Traffic Light Control for Intelligent Transportation System. *IEEE Trans. Intell. Transp. Syst.* **2020**, 1–10. [CrossRef]
7. Barbosa, R.C.; Ayub, M.S.; Rosa, R.L.; Rodríguez, D.Z.; Wuttisittikulkij, L. Lightweight PVIDNet: A priority vehicles detection network model based on deep learning for intelligent traffic lights. *Sensors* **2020**, *20*, 6218. [CrossRef]
8. Veres, M.; Moussa, M. Deep Learning for Intelligent Transportation Systems: A Survey of Emerging Trends. *IEEE Trans. Intell. Transp. Syst.* **2020**, *21*, 3152–3168. [CrossRef]
9. He, P.; Wu, A.; Huang, X.; Scott, J.; Rangarajan, A.; Ranka, S. Truck and Trailer Classification With Deep Learning Based Geometric Features. *IEEE Trans. Intell. Transp. Syst.* **2020**, 1–10. [CrossRef]
10. Lasmar, E.L.; de Paula, F.O.; Rosa, R.L.; Abrahão, J.I.; Rodríguez, D.Z. RsRS: Ridesharing Recommendation System Based on Social Networks to Improve the User's QoE. *IEEE Trans. Intell. Transp. Syst.* **2019**, *20*, 4728–4740. [CrossRef]
11. Huh, J.H.; Seo, K. Artificial Intelligence Shoe Cabinet Using Deep Learning for Smart Home. In *Advanced Multimedia and Ubiquitous Engineering*; Park, J.J., Loia, V., Choo, K.K.R., Yi, G., Eds.; Springer: Singapore, 2019; pp. 825–834.
12. Rosa, R.L.; Rodríguez, D.Z.; Bressan, G. SentiMeter-Br: A new social web analysis metric to discover consumers' sentiment. In Proceedings of the IEEE International Symposium on Consumer Electronics (ISCE), Hsinchu, Taiwan, 3–6 June 2013; pp. 153–154. [CrossRef]
13. Zinemanas, P.; Rocamora, M.; Miron, M.; Font, F.; Serra, X. An Interpretable Deep Learning Model for Automatic Sound Classification. *Electronics* **2021**, *10*, 850. [CrossRef]
14. Chen, Z.; Ma, G.; Jiang, Y.; Wang, B.; Soleimani, M. Application of Deep Neural Network to the Reconstruction of Two-Phase Material Imaging by Capacitively Coupled Electrical Resistance Tomography. *Electronics* **2021**, *10*, 1058. [CrossRef]

15. Akhand, M.A.H.; Roy, S.; Siddique, N.; Kamal, M.A.S.; Shimamura, T. Facial Emotion Recognition Using Transfer Learning in the Deep CNN. *Electronics* **2021**, *10*, 1036. [CrossRef]
16. Guimarães, R.; Rodríguez, D.Z.; Rosa, R.L.; Bressan, G. Recommendation system using sentiment analysis considering the polarity of the adverb. In Proceedings of the IEEE International Symposium on Consumer Electronics (ISCE), Sao Paulo, Brazil, 28–30 September 2016; pp. 71–72. [CrossRef]
17. Azar, A.T.; Koubaa, A.; Ali Mohamed, N.; Ibrahim, H.A.; Ibrahim, Z.F.; Kazim, M.; Ammar, A.; Benjdira, B.; Khamis, A.M.; Hameed, I.A.; et al. Drone Deep Reinforcement Learning: A Review. *Electronics* **2021**, *10*, 999. [CrossRef]
18. Wang, X.; Chan, K.C.; Yu, K.; Dong, C.; Loy, C.C. EDVR: Video Restoration with Enhanced Deformable Convolutional Networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Long Beach, CA, USA, 16–17 June 2019; pp. 1954–1963. [CrossRef]
19. Militani, D.R.; de Moraes, H.P.; Rosa, R.L.; Wuttisittikulkij, L.; Ramírez, M.A.; Rodríguez, D.Z. Enhanced Routing Algorithm Based on Reinforcement Machine Learning—A Case of VoIP Service. *Sensors* **2021**, *21*, 504. [CrossRef]
20. Kim, H.M.; Kim, M.S.; Lee, G.J.; Yoo, Y.J.; Song, Y.M. Large area fabrication of engineered microlens array with low sag height for light-field imaging. *Opt. Express* **2019**, *27*, 4435–4444. [CrossRef]
21. Perra, C. Assessing the Quality of Experience in Viewing Rendered Decompressed Light Fields. *Multimed. Tools Appl.* **2018**, *77*, 21771–21790. [CrossRef]
22. Kovács, P.T.; Bregović, R.; Boev, A.; Barsi, A.; Gotchev, A. Quantifying Spatial and Angular Resolution of Light-Field 3-D Displays. *IEEE J. Sel. Top. Signal Process.* **2017**, *11*, 1213–1222. [CrossRef]
23. Kalantari, N.K.; Wang, T.C.; Ramamoorthi, R. Learning-Based View Synthesis for Light Field Cameras. *ACM Trans. Graph.* **2016**, *35*, 1–10. [CrossRef]
24. Meng, N.; So, H.K.; Sun, X.; Lam, E. High-dimensional Dense Residual Convolutional Neural Network for Light Field Reconstruction. *IEEE Trans. Pattern Anal. Mach. Intell.* **2019**, 43, 873–886. [CrossRef]
25. Affonso, E.T.; Rodríguez, D.Z.; Rosa, R.L.; Andrade, T.; Bressan, G. Voice quality assessment in mobile devices considering different fading models. In Proceedings of the 2016 IEEE International Symposium on Consumer Electronics (ISCE), Sao Paulo, Brazil, 28–30 September 2016; pp. 21–22. [CrossRef]
26. Rosa, R.L.; Schwartz, G.M.; Ruggiero, W.V.; Rodríguez, D.Z. A Knowledge-Based Recommendation System That Includes Sentiment Analysis and Deep Learning. *IEEE Trans. Ind. Inform.* **2019**, *15*, 2124–2135. [CrossRef]
27. Affonso, E.T.; Nunes, R.D.; Rosa, R.L.; Pivaro, G.F.; Rodríguez, D.Z. Speech Quality Assessment in Wireless VoIP Communication Using Deep Belief Network. *IEEE Access* **2018**, *6*, 77022–77032. [CrossRef]
28. Zhang, L.; Wu, J.; Fan, Y.; Gao, H.; Shao, Y. An Efficient Building Extraction Method from High Spatial Resolution Remote Sensing Images Based on Improved Mask R-CNN. *Sensors* **2020**, *20*, 1465. [CrossRef]
29. Wang, T.; Efros, A.A.; Ramamoorthi, R. Depth Estimation with Occlusion Modeling Using Light-Field Cameras. *IEEE Trans. Pattern Anal. Mach. Intell.* **2016**, *38*, 2170–2181. [CrossRef] [PubMed]
30. Fei, L.; Hou, G.; Sun, Z.; Tan, T. High Quality Depth Map Estimation of Object Surface from Light Field Images. *Neurocomputing* **2017**, *252*, 3–16. [CrossRef]
31. Wang, Y.; Zhang, J. Reconstruction of compressively sampled light field by using tensor dictionaries. *Multimed. Tools Appl.* **2020**, *79*, 20449–20460. [CrossRef]
32. Garg, R.; Bg, V.K.; Carneiro, G.; Reid, I. Unsupervised CNN for Single View Depth Estimation: Geometry to the Rescue. In *Proceedings of the Computer Vision—ECCV 2016, Amsterdam, The Netherlands, 11–14 October 2016*; Leibe, B., Matas, J., Sebe, N., Welling, M., Eds.; Springer International Publishing: Cham, Switzerland, 2016; pp. 740–756.
33. 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 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 1638–1646.
34. Wu, G.; Liu, Y.; Fang, L.; Dai, Q.; Chai, T. Light Field Reconstruction Using Convolutional Network on EPI and Extended Applications. *IEEE Trans. Pattern Anal. Mach. Intell.* **2019**, *41*, 1681–1694. [CrossRef] [PubMed]
35. Xiaoping, L.; Zhenjun, T.; Zhenjun, T.; Xiaolan, X.; Xianquan, Z. Robust and fast image hashing with two-dimensional PCA. *Multimed. Syst.* **2020**, *1*, 4435–4444. [CrossRef]
36. Veerasamy, B.; Annadurai, S. Video compression using hybrid hexagon search and teaching–learning-based optimization technique for 3D reconstruction. *Multimed. Syst.* **2020**, *1*, 1–15. [CrossRef]
37. Xu, Z.; Zhang, Q.; Cheng, S. Multilevel active registration for kinect human body scans: From low quality to high quality. *Multimed. Syst.* **2018**, *24*, 257–270. [CrossRef]
38. Shah, N.M.H.; Junaid, M.; Faseeh, N.M.; Shin, D.R. A Multi-blocked Image Classifier for Deep Learning. *Mehran Univ. Res. J. Eng. Technol.* **2020**, *39*, 583–594. [CrossRef]
39. Tong, C.; Liang, B.; Su, Q.; Yu, M.; Hu, J.; Bashir, A.K.; Zheng, Z. Pulmonary Nodule Classification Based on Heterogeneous Features Learning. *IEEE J. Sel. Areas Commun.* **2020**, *39*, 574–581. [CrossRef]
40. Ashraf, R.; Afzal, S.; Rehman, A.U.; Gul, S.; Baber, J.; Bakhtyar, M.; Mehmood, I.; Song, O.Y.; Maqsood, M. Region-of-Interest Based Transfer Learning Assisted Framework for Skin Cancer Detection. *IEEE Access* **2020**, *8*, 147858–147871. [CrossRef]
41. Saadi, M.; Ahmad, T.; Kamran Saleem, M.; Wuttisittikulkij, L. Visible light communication–an architectural perspective on the applications and data rate improvement strategies. *Trans. Emerg. Telecommun. Technol.* **2019**, *30*, e3436. [CrossRef]

42. Fu, L.; Ren, C.; He, X.; Wu, X.; Wang, Z. Single Remote Sensing Image Super-Resolution with an Adaptive Joint Constraint Model. *Sensors* **2020**, *20*, 1276. [CrossRef] [PubMed]
43. 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]
44. Rossi, M.; Frossard, P. Geometry-Consistent Light Field Super-Resolution via Graph-Based Regularization. *IEEE Trans. Image Process.* **2018**, *27*, 4207–4218. [CrossRef]
45. Ghassab, V.K.; Bouguila, N. Light Field Super-Resolution Using Edge-Preserved Graph-Based Regularization. *IEEE Trans. Multimed.* **2020**, *22*, 1447–1457. [CrossRef]
46. 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.
47. 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]
48. 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 (CVPR), Long Beach, CA, USA, 15–20 June 2019.
49. 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]
50. 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.* **2019**, *28*, 2319–2330. [CrossRef]
51. Saadi, M.; Saeed, Z.; Ahmad, T.; Saleem, M.K.; Wuttisittikulkij, L. Visible light-based indoor localization using k-means clustering and linear regression. *Trans. Emerg. Telecommun. Technol.* **2019**, *30*, e3480. [CrossRef]
52. Qiao, F.; Wu, J.; Li, J.; Bashir, A.K.; Mumtaz, S.; Tariq, U. Trustworthy edge storage orchestration in intelligent transportation systems using reinforcement learning. *IEEE Trans. Intell. Transp. Syst.* **2020**, 1–14. [CrossRef]
53. Ji, B.; Chen, Z.; Mumtaz, S.; Liu, J.; Zhang, Y.; Zhu, J.; Li, C. SWIPT Enabled Intelligent Transportation Systems with Advanced Sensing Fusion. *IEEE Sens. J.* **2020**, *1*. [CrossRef]
54. Noomwongs, N.; Bajpai, A.; Phutthaburee, P.; Wongpiya, L.; Skulthai, A.; Maung, T.Z.B.; Myint, Y.M.; Ullah, I.; Wuttisittikulkij, L.; Saadi, M. Design and Testing of Autonomous Steering System Implemented on a Toyota Ha: mo. In Proceedings of the 2020 International Conference on Electronics, Information, and Communication (ICEIC), Barcelona, Spain, 19–22 January 2020; IEEE: New York, NY, USA, 2020; pp. 1–5.
55. Du, G.; Wang, Z.; Gao, B.; Mumtaz, S.; Abualnaja, K.M.; Du, C. A Convolution Bidirectional Long Short-Term Memory Neural Network for Driver Emotion Recognition. *IEEE Trans. Intell. Transp. Syst.* **2020**, 1–9. [CrossRef]
56. Khan, M.Z.; Harous, S.; Hassan, S.U.; Khan, M.U.G.; Iqbal, R.; Mumtaz, S. Deep unified model for face recognition based on convolution neural network and edge computing. *IEEE Access* **2019**, *7*, 72622–72633. [CrossRef]
57. Lin, Z.; Shum, H.Y. A Geometric Analysis of Light Field Rendering. *Int. J. Comput. Vis.* **2004**, *58*, 121–138. [CrossRef]
58. Zhao, Q.; Dai, F.; Lv, J.; Ma, Y.; Zhang, Y. Panoramic Light Field From Hand-Held Video and Its Sampling for Real-Time Rendering. *IEEE Trans. Circuits Syst. Video Technol.* **2020**, *30*, 1011–1021. [CrossRef]
59. Suzuki, T.; Takahashi, K.; Fujii, T. Disparity estimation from light fields using sheared EPI analysis. In Proceedings of the 2016 IEEE International Conference on Image Processing (ICIP), Phoenix, AZ, USA, 25–28 September 2016; pp. 1444–1448.
60. 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.
61. Ha, I.Y.; Wilms, M.; Heinrich, M. Semantically Guided Large Deformation Estimation with Deep Networks. *Sensors* **2020**, *20*, 1392. [CrossRef] [PubMed]
62. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
63. Huang, G.; Liu, Z.; Van Der Maaten, L.; Weinberger, K.Q. Densely Connected Convolutional Networks. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 2261–2269.
64. Lu, Z.; Yeung, H.; Qu, Q.; Chung, Y.; Chen, X.; Chen, Z. Improved image classification with 4D light-field and interleaved convolutional neural network. *Multimed. Tools Appl.* **2019**, *78*, 29211–29227. [CrossRef]
65. Mendonça, R.V.; Teodoro, A.A.M.; Rosa, R.L.; Saadi, M.; Melgarejo, D.C.; Nardelli, P.H.J.; Rodríguez, D.Z. Intrusion Detection System Based on Fast Hierarchical Deep Convolutional Neural Network. *IEEE Access* **2021**, *9*, 61024–61034. [CrossRef]
66. Terra Vieira, S.; Lopes Rosa, R.; Zegarra Rodríguez, D.; Arjona Ramírez, M.; Saadi, M.; Wuttisittikulkij, L. Q-Meter: Quality Monitoring System for Telecommunication Services Based on Sentiment Analysis Using Deep Learning. *Sensors* **2021**, *21*, 1880. [CrossRef] [PubMed]
67. Dong, C.; Loy, C.C.; He, K.; Tang, X. Image Super-Resolution Using Deep Convolutional Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2016**, *38*, 295–307. [CrossRef]
68. Dong, C.; Loy, C.C.; Tang, X. *Accelerating the Super-Resolution Convolutional Neural Network*; Springer: Cham, Switzerland, 2016; Volume 9906, pp. 391–407. [CrossRef]
69. Kim, J.; Lee, J.K.; Lee, K.M. Deeply-Recursive Convolutional Network for Image Super-Resolution. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 1637–1645. [CrossRef]

70. Lai, W.; Huang, J.; Ahuja, N.; Yang, M. Deep Laplacian Pyramid Networks for Fast and Accurate Super-Resolution. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 5835–5843.

71. Hu, X.; Mu, H.; Zhang, X.; Wang, Z.; Tan, T.; Sun, J. Meta-SR: A Magnification-Arbitrary Network for 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. 1575–1584.

72. Yoon, Y.; Jeon, H.; Yoo, D.; Lee, J.; Kweon, I.S. Light-Field Image Super-Resolution Using Convolutional Neural Network. *IEEE Signal Process. Lett.* **2017**, *24*, 848–852. [CrossRef]

73. Tian, Y.; Zeng, H.; Hou, J.; Chen, J.; Ma, K. Light Field Image Quality Assessment via the Light Field Coherence. *IEEE Trans. Image Process.* **2020**, *29*, 7945–7956. [CrossRef]

74. Lim, B.; Son, S.; Kim, H.; Nah, S.; Lee, K.M. Enhanced Deep Residual Networks for Single Image Super-Resolution. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Honolulu, HI, USA, 21–26 July 2017; pp. 1132–1140.

75. Wang, Y.; Wang, L.; Yang, J.; An, W.; Yu, J.; Guo, Y. Spatial-Angular Interaction for Light Field Image Super-Resolution. *arXiv* **2020**, arXiv:1912.07849.

76. 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 (CVPR), Seattle, WA, USA, 13–19 June 2020.

77. Dai, J.; Qi, H.; Xiong, Y.; Li, Y.; Zhang, G.; Hu, H.; Wei, Y. Deformable Convolutional Networks. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 Ocotber 2017.

78. Tian, Y.; Zhang, Y.; Fu, Y.; Xu, C. TDAN: Temporally-Deformable Alignment Network for Video Super-Resolution. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020.

79. Xiang, X.; Tian, Y.; Zhang, Y.; Fu, Y.; Allebach, J.P.; Xu, C. Zooming Slow-Mo: Fast and Accurate One-Stage Space-Time Video Super-Resolution. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020.

80. Le Pendu, M.; Jiang, X.; Guillemot, C. Light Field Inpainting Propagation via Low Rank Matrix Completion. *IEEE Trans. Image Process.* **2018**, *27*, 1981–1993. [CrossRef]

81. Honauer, K.; Johannsen, O.; Kondermann, D.; Goldluecke, B. A Dataset and Evaluation Methodology for Depth Estimation on 4D Light Fields. In *Proceedings of the Computer Vision—ACCV 2016, Taipei, Taiwan, 20–24 November 2016*; Lai, S.H., Lepetit, V., Nishino, K., Sato, Y., Eds.; Springer International Publishing: Cham, Switzerland, 2017; pp. 19–34.

82. Rerabek, M.; Ebrahimi, T. New Light Field Image Dataset. In Proceedings of the 8th International Workshop on Quality of Multimedia Experience (QoMEX), Lisbon, Portugal, 6–8 June 2016.

83. Wanner, S.; Meister, S.; Goldluecke, B. Datasets and Benchmarks for Densely Sampled 4D Light Fields. In Proceedings of the 18th International Workshop on Vision, Modeling and Visualization (VMV 2013), Lugano, Switzerland, 11–13 September 2013; pp. 225–226.

84. Zhang, Y.; Li, K.; Li, K.; Wang, L.; Zhong, B.; Fu, Y. Image Super-Resolution Using Very Deep Residual Channel Attention Networks. In Proceedings of the Computer Vision—ECCV 2018—15th European Conference, Munich, Germany, 8–14 September 2018; Proceedings, Part VII; Lecture Notes in Computer Science; Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y., Eds.; Springer: Berlin, Germany, 2018; Volume 11211, pp. 294–310. [CrossRef]

85. Dai, T.; Cai, J.; Zhang, Y.; Xia, S.T.; Zhang, L. Second-Order Attention Network for Single Image Super-Resolution. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019.