IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received May 21, 2020, accepted June 20, 2020, date of publication June 23, 2020, date of current version July 2, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3004477

LightGAN: A Deep Generative Model for Light Field Reconstruction

NAN MENG, ZHOU GE, TIANJIAO ZENG, AND EDMUND Y. LAM, (Fellow, IEEE)
Department of Electrical and Electronic Engineering, The University of Hong Kong, Hong Kong

Corresponding author: Edmund Y. Lam (elam@eee.hku.hk)

This work was supported in part by the Research Grants Council of Hong Kong under Grant GRF 17203217, Grant 17201818, and Grant 17200019, and in part by the University of Hong Kong under Grant 104005009 and Grant 104005438.

ABSTRACT A light field image captured by a plenoptic camera can be considered a sampling of light distribution within a given space. However, with the limited pixel count of the sensor, the acquisition of a high-resolution sample often comes at the expense of losing parallax information. In this work, we present a learning-based generative framework to overcome such tradeoff by directly simulating the light field distribution. An important module of our model is the high-dimensional residual block, which fully exploits the spatio-angular information. By directly learning the distribution, our approach can generate both high-quality sub-aperture images and densely-sampled light fields. Experimental results on both real-world and synthetic datasets demonstrate that the proposed method outperforms other state-of-the-art approaches and achieves visually more realistic results.

INDEX TERMS Light field reconstruction, view synthesis, 4D convolution, high-dimension residual block, generative adversarial networks, deep learning, computational imaging.

I. INTRODUCTION

In computer vision and three-dimensional imaging, light field imaging has generated a lot of interest due to the designs of various capturing systems [1], [2]. Compared to conventional cameras, a light field camera (also known as a plenoptic camera) allows one to capture both intensity values and directions of light rays from real-world scenes. The additional information enables many applications, such as image refocusing [3], depth estimation [4], [5], and novel view generation [6], [7]. However, there is an inherent tradeoff between spatial and angular resolutions. The generally lower spatial resolution of the light field image poses great challenges in exploiting the advantages brought from additional angular sampling [8], [9].

Taking advantage of the parallax between two neighboring views, the captured light field scenes preserve high correlations among the sub-aperture images (SAIs). Others addressing the light field super-resolution problem generally regard geometry properties as the reconstruction priors, and warp the neighboring views to the target view [10], [11]. The performance of these methods depends on accurate geometric information of the scene as priors. However, approaches for depth estimation have difficulties in providing accurate depth estimation for pixel warping. Errors in this process give rise to artifacts such as tearing and ghosting.

To mitigate the dependency on explicit depth or disparity information, many alternative approaches are based on sampling and consecutive reconstruction of the plenoptic function [12], [13]. Instead of using the disparity as auxiliary information, they consider each pixel of the given SAI as a sample of the light field distribution function. Recently, deep learning has been proved to be a powerful technique in a wide range of applications [14], [15]. With the availability of the light field dataset [16], methods based on the convolutional neural networks (CNNs) have been successfully applied to light field super-resolution [17], [18]. Yoon *et al.* [19] establish the first deep learning framework LFCNN for both spatial and angular super-resolution but do not exploit the correlation among adjacent views. Wang *et al.* [20] regard the light field as an image sequence and introduce the bidirectional recurrent convolutional neural network to approximate the correspondences of neighboring images. However, the image sequence assumption reduces the complexity of light field

The associate editor coordinating the review of this manuscript and approving it for publication was Shuhan Shen.

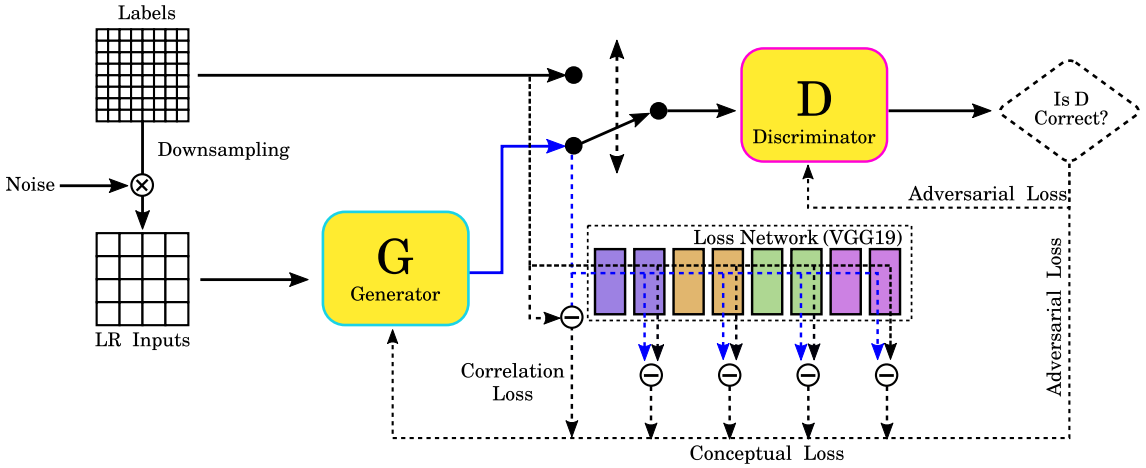


FIGURE 1. Overview of the proposed LightGAN. The generator takes the low-resolution light fields as inputs and generates the super-resolved counterparts, which are further judged by the discriminator together with the high-resolution labels. The results of the discriminator are used to obtain the adversarial loss, and they are combined with the conceptual loss and the correlation loss to supervise the generator training process.

angular correlation and changes the relations among the surrounding SAIs, which consequently limits the reconstruction results.

Given the inherent geometric properties of light field data, reconstruction algorithms should involve information from both spatial and angular dimensions. However, existing methods struggle to handle the uncertainty in recovering lost high-frequency spatial details while preserving angular correlation. In this paper, we propose a generative model to effectively address the light field super-resolution problem. The generative adversarial networks (GANs) are well known for their powerful capacity in generating plausible-looking natural images with high perceptual quality [21], [22]. Considering such benefits, we incorporate the high-dimensional convolution (HDC) layers into the GAN framework to learn the high correlations among the neighboring light field views. The proposed model is tailored to the structural property of the light field, which therefore is named LightGAN.

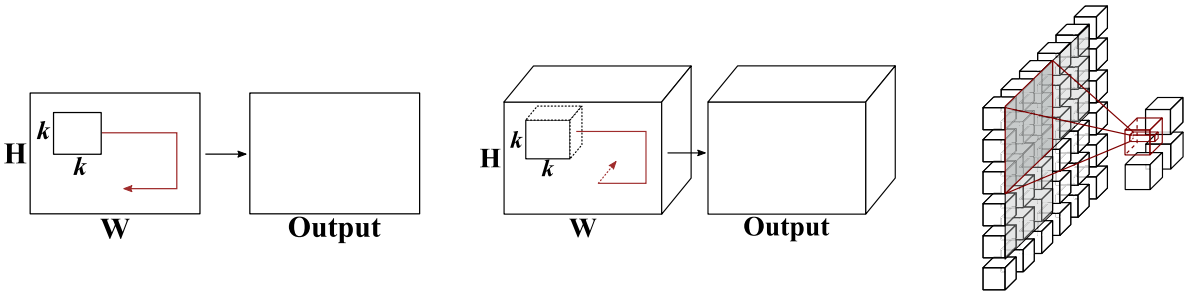
Fig. 1 presents an overview of the proposed LightGAN framework. Our model is trained by minimizing the combination of the conceptual and correlation loss together with the adversarial loss. Unlike other deep learning frameworks, our model is particularly designed for light field data to fully exploit the complete spatio-angular correlations, and can address the low-resolution problem in both spatial and angular dimensions.

II. RELATED WORK

Existing light field reconstruction methods can be mainly divided into two different categories. Approaches of the first class require an accurate estimation of geometry and expensive computations [23], [24]. For instance, Bishop and Favaro [25] design a Bayesian framework to recover the pixels in the super-resolved light field by efficiently exploiting the geometric structure. Mitra and Veeraraghavan [3] take a patch-based approach, and apply a fast subspace projection technique to generate the disparity map and wrap the disparity information into light field images by the Gaussian mixture model. Later, Wang *et al.* [26] propose an algorithm that relaxes the dependency of depth information in projection-based models via a re-defined mapping function between the disparity of certain pixel and its shearing shift. The depth-based methods tend to fail in occlusion regions or non-Lambertain surfaces where the depth information is difficult to obtain.

In addition to the depth-dependent approaches, some alternative attempts focus on the projection and resampling of light field data [27], [28]. Liang and Ramamoorthi [29] provide a theoretical analysis on the resolution limit of light field and demonstrate that the lenslet-based cameras can achieve spatial resolution above the microlens resolution. Meanwhile, Cho *et al.* [30] illustrate the calibration procedure of a raw light field image and they further propose a dictionary-learning interpolation method for light field reconstruction.

The second category makes the use of deep learning frameworks. With the development of CNNs, backed by having more and more light field datasets [31], [32], some learning-based methods have been proposed recently and show promising performance [33], [34]. Kalantari *et al.* [35] introduce the CNNs to traditional pipeline and produce plausible images. Zhang *et al.* [36] adopt a branched residual network for spatial super-resolution. Different branches learn the relations of SAIs in different directions. All the learned features are finally combined together to generate the enhanced light field. Farrugia *et al.* [37], on the other hand, exploit a dictionary learning-based method which learns the mapping between the low-resolution and high-resolution patches. These methods exploit the angular correlations, but not specially tailored for the light field structure.



(a) 2D Convolution (b) 3D Convolution (c) 4D Convolution

FIGURE 2. Illustration of 2D, 3D and 4D convolution operations. a) Illustration of applying the 2D convolution on the 2D data, which results in an image. b) Illustration of applying the 3D convolution on the 3D data, which results in a volume. c) Illustration of applying the 4D convolution on the 4D data, which results in a 4D tesseract.

Recently, Yeung *et al.* [38] and Wang *et al.* [39] introduce the 4D convolution to drive the network to learn the complex correlations in angular dimensions. Compared with previous models, the 4D convolution is more fit for light field data and exhibits powerful ability in learning the high-dimensional correlations, allowing the framework to be trained end-to-end. Nevertheless, both of their models only focus on view synthesis. By contrast, this paper presents a framework that is based on HDC as well, but is able to enhance both spatial and angular resolution.

III. METHODOLOGY

A. PROBLEM FORMULATION

Given a low-resolution light field $I^L(x, y, u, v)$ at the resolution of (X, Y, U, V), the goal of light field super-resolution is to reconstruct its corresponding high-resolution counterpart $I^H(x, y, u, v)$ at the resolution of $(\gamma_s X, \gamma_s Y, \gamma_a U, \gamma_a V)$ which can be formulated as

$$I^L(x, y, u, v) \xrightarrow{g} I^H(x, y, u, v), \tag{1}$$

where γ_s and γ_a correspond to the spatial and angular upscaling factors, respectively. (X, Y) denotes the resolution in the spatial dimensions, while (U, V) represents the resolution in the angular dimensions. The function $g(\cdot)$ in Eq. 1 denotes the recovery mapping between I^L and I^H, which is approximated using a generative network in this study.

The objective of the GAN is to learn a distribution that resembles the real data distribution [40]. Such a property is desirable for our task aiming at simulating the light field distribution based on a few samples. The algorithm trains a discriminator D to maximize the probability of correctly recognizing the label and the enhanced light field, i.e. $\log(D(I^H))$. Simultaneously, it also trains a generator G to minimize the loss, given by $\log(1 - D(G(I^L)))$. Therefore, the framework is trained by optimizing the value function $V(D, G)$, i.e.,

$$\min_{\theta_G}\max_{\theta_D} V(D, G) = \mathbb{E}_{I^H\sim\pi(I^H)}\left[\log D\left(I^H\right)\right] + \mathbb{E}_{I^L\sim\pi_G(I^L)}\left[\log\left(1-D\left(G\left(I^L\right)\right)\right)\right], \tag{2}$$

where θ_G and θ_D represent the parameters of G and D, respectively. $\mathbb{E}(\cdot)$ is the expectation function. During the optimization process, the generator G learns the mapping $g(\cdot)$ in Eq. 1 and outputs a high-resolution light field that resembles the original one. $\pi(\cdot)$ stands for the light field distribution described in the training samples, while $\pi_G(\cdot)$ is the distribution of the inputs. We assume that the training set is well-sampled from the real scenes. The idea behind this formulation is that it builds up the generator G with the goal of fooling a discriminator D that is trained to distinguish the reconstructed and real scenes. This encourages the solutions residing in the space of light field images, leading to results that preserve the geometry information of the scene.

Our ultimate goal is to acquire a mapping function that estimates the real light field distribution and generates the high-resolution counterpart from a given low-resolution input. To achieve this, we train a generative network parametrized by $\Theta = \{\theta_i\}$, where $i = 1, \ldots, L$. Here, θ_i denotes the weight and bias parameters of the i^{th} layer in the L-layer network. Therefore, for the training labels $\{I_n^H\}$ and corresponding low-resolution inputs $\{I_n^L\}$, where $n = 1, \ldots, N$, the mapping parameters Θ is obtained by optimizing a loss function $\mathcal{L}(\cdot)$, such that

$$\Theta^* = \arg\min_{\Theta} \frac{1}{N}\sum_{n=1}^{N} \mathcal{L}\left(G(I_n^L; \Theta), I_n^H\right). \tag{3}$$

Specifically in this work, $\mathcal{L}(\cdot)$ is designed as a weighted combination of several loss components, which encourages our proposed model to learn the spatial details while preserving geometry properties.

B. HIGH-DIMENSIONAL CONVOLUTION

Compared with 2D or 3D convolution, the HDC is operated in the 4D space, leading to its strong capacity to fully calculate the spatio-angular redundancy. In 2D CNN models, the intrinsic limitation makes it hard to handle the problem with more than two dimensions, as illustrated in Fig. 2. Hence, most existing learning-based methods apply CNNs on either SAIs [19], [35] or EPI images [41] to learn the relations between neighboring views. Another assumption for

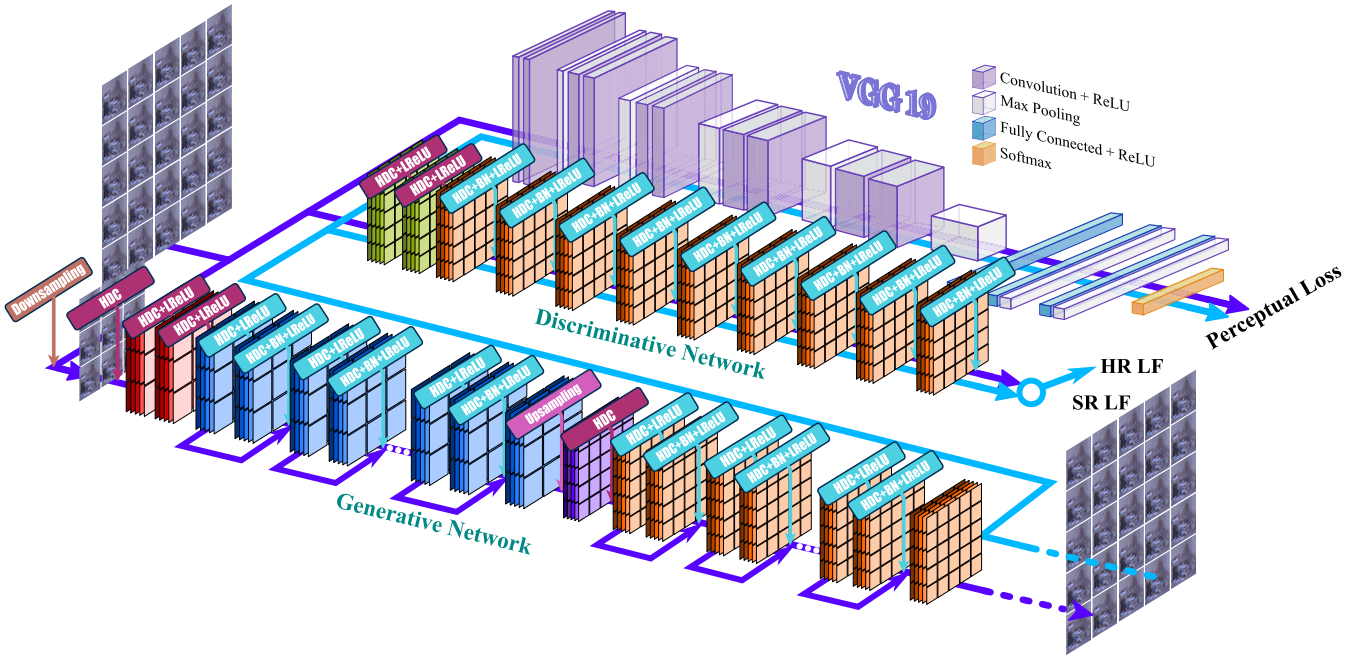


FIGURE 3. Detailed framework for the proposed LightGAN. The generator is established by densely-connected multiple HDRB and HDC layers, while the discriminator is constructed using a simple linear structure.

light field processing is to consider the 4D data as image sequences. Several methods are based on this assumption and apply the approaches from video processing to handle the sequences [42], [43]. However, such an assumption omits the relations between the spatial and angular coordinates. The result is that these approaches have trouble approximating the correct EPIs of occluded or non-Lambertian regions, which therefore causes ghosting or tearing artifacts near the object boundaries. By contrast, the HDC layer shows its potential in processing the light field image for multiple applications, such as material recognition [16], view synthesis [38], and super-resolution [34]. These achievements are owed greatly to its ability to extract spatial representations preserving the angular correlations. As a consequence, by incorporating the HDC layer, the proposed model can make full use of the structural information to simulate the original light field distribution.

C. NETWORK ARCHITECTURE

The GAN-based model maintains the ability to drive the reconstruction towards image manifold [21]. For light field, more importantly, the reconstruction should guarantee the epipolar property of data. Therefore, in the proposed network, we not only ensure that the generator can process high-dimensional data, but also allow the discriminator to witness the entire light field image in making the judgment. The entire architecture of the proposed GAN is presented in Fig. 3. Both the generator and the discriminator process the light field data directly in 4D space. The generative network is established using multiple high-dimensional convolutional residual blocks (HDRB). As depicted in Fig. 4, each HDRB

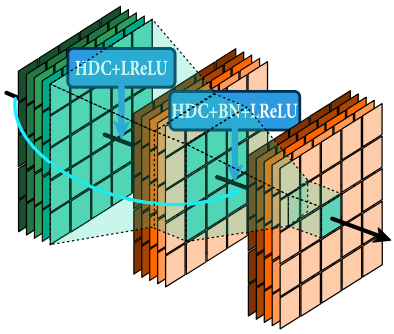


FIGURE 4. Illustration of a high-dimensional residual block (HDRB).

is composed of two HDC layers [18] with the 3×3 angular receptive field. With such a structure, the angular receptive field of every HDRB will cover the entire 5×5 viewpoints, allowing the module to learn the complete spatio-angular structure and redundancy of a light field.

Furthermore, inspired by an earlier work on image recognition [44], we design the HDRB module and use the Leaky Rectified Linear Unit (LReLU) [45] as the activation function The spatial upscaling operation is performed by a subpixel convolution layer [46] while the angular pixel is upscaled using the linear interpolation method. The output of the generator network, together with real-world high-resolution images, are fed into the discriminator network for training. As Fig. 3 shows, the whole structure of the discriminator network contains several HDC convolutional layers with increasing filter depth from 64 to 512, which is the same as the network proposed by Visual Geometry Group (VGG team) [47].

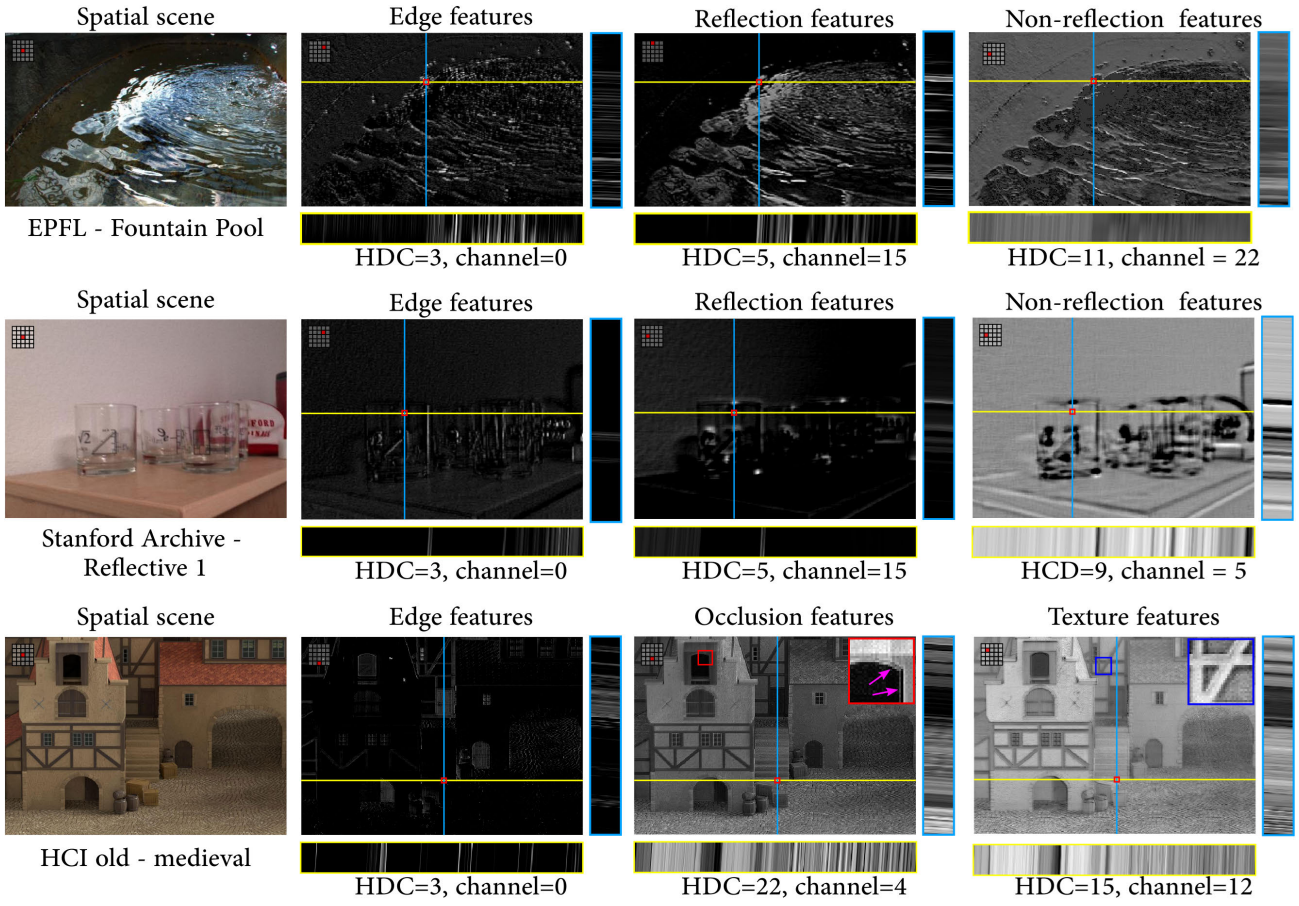


FIGURE 5. Visualization of the geometric features extracted from different HDC layers of the proposed LightGAN.

However, we replace the fully-connected layers with a mean layer (calculating the average output of the final convolution layer), which also supports the generator to converge to a good solution according to the results, but dramatically reduces the parameters compared with densely connected layers.

D. HIGH-DIMENSIONAL FEATURE MAPS

As discussed in Section III-B, the HDC layer is able to process the entire spatio-angular information of light field data and extract the features preserving the angular correlations. To figure out such a property, we dig into the learned high-dimensional features and visualize the spatial appearances and EPI patterns of these feature maps in Fig. 5. In the figure, we compare the feature maps of three representative light fields including two real-world scenes and one synthetic scene. The first column gives the center view of each light field. According to their view images, the selected samples contain multiple types of surfaces, including the reflection surfaces (i.e. bright regions in the "water" surface), transmission surfaces (i.e. dark regions in the "water" surface and the "glass" regions), and Lambertain surfaces (i.e. "wall" and "floor" regions, etc). The rest three columns exhibit the spatial appearances and EPI patterns of the learned features. The bright regions of the feature maps denote the places with high activation. For example, the feature maps in the 2nd column have higher activation near the object border. This indicates that the features extracted from the 3rd HDC layer contain edge features. Likewise, other layers can extract the reflection features, occlusion features, texture features, etc. Although the spatial appearances are diverse, one common property of these high-dimensional features is they preserve the structural information of light field, which can be demonstrated by the feature EPIs presented in Fig. 5. The EPI patterns of feature maps are close to the light field EPIs. Therefore, the proposed generator, to some extend, "remembers" the structural information in its learned features.

E. LOSS FUNCTION

The definition of our loss function is critical for the generator network performance on light field reconstruction. We design a novel spatio-angular loss function ℓ_{SA} to evaluate the restoration results of light field. Such a spatio-angular loss is formulated as a weighted combination of a spatial conceptual loss ℓ_S, an angular correlation loss ℓ_A, and a generative adversarial loss ℓ_G, i.e.,

$$\ell_{\mathrm{SA}} = \alpha \cdot \ell_{\mathrm{S}} + \beta \cdot \ell_{\mathrm{A}} + \gamma \cdot \ell_{\mathrm{G}}, \tag{4}$$

where the scalars α, β and γ denote the weights of each loss term. The first term ℓ_S describes the conceptual differences between each labels of SAI and reconstructed SAI. The second term ℓ_A is defined based on the mean square error (MSE) loss, which measures the ℓ_2 differences between the entire reconstruction and corresponding label. The third term ℓ_G is the inherent loss term of the GAN framework, which is used to measure the stability of the competition between the generator and the discriminator. These three loss terms are further described below.

1) SPATIAL CONCEPTUAL LOSS

Inspired by the work of Ledig *et al.* on single image super-resolution [21], we introduce the perceptual loss to describe the spatial difference between each pair of corresponding SAIs. Such loss is derived by applying the VGG loss ℓ_{VGG} on each SAI. The motivation of employing the VGG loss is based on the fact that the SAI can be easily perceived by humans [48]. This reflects that every SAI contains the core representations similar to a natural image. Therefore, the expression of spatial conceptual loss term can be formulated as

$$\ell_S = \frac{1}{UV}\sum_{u=1}^{U}\sum_{v=1}^{V}\left(\phi\left(I^H_{u,v}\right) - \phi\left(G\left(I^L_{u,v}\right)\right)\right)^2, \tag{5}$$

where $\phi(\cdot)$ denotes the mapping described in [21]. For clarity, we define two notations $I^H_{u,v} = I^H(\cdot,\cdot,u,v)$ and $I^L_{u,v} = I^L(\cdot,\cdot,u,v)$ to represent the SAIs (with angular coordinate (u,v)) of I^H and I^L, respectively. As formulated in Eq. 5, the conceptual loss is calculated on each pair of SAIs. One is from the ground truth, and the other is from the reconstructed light field.

2) ANGULAR CORRELATION LOSS

The angular correlation loss is defined based on the MSE, which calculates the differences between the label EPIs and output EPIs. Such a loss term can be formalized as

$$\ell_A = \frac{1}{XU}\sum_{x=1}^{X}\sum_{u=1}^{U}\left(E^H_{x,u} - G(E^L_{x,u})\right)^2. \tag{6}$$

In this equation, we use the EPIs acquired by fixing a spatial coordinate and an angular coordinate to calculate the correlation loss. In other words, $E^H_{x,u} = I^H(x,\cdot,u,\cdot)$ and $E^L_{x,u} = I^L(x,\cdot,u,\cdot)$. The EPI pattern describes the parallax of each pixel in different images. As a result, the angular correlation loss encourages the EPIs of generated light field to be close to the ones of the original light field.

3) GENERATIVE ADVERSARIAL LOSS

The last loss term is the inherent generative adversarial loss. For each input I^L, the loss is computed as

$$\ell_G = \log\left(1 - D\left(G\left(I^L\right)\right)\right). \tag{7}$$

TABLE 1. Quantitative Performance of the proposed network trained using different loss terms on HCI new testset for spatial 4× task.

Settings	Spatial loss ℓ_S	Angular loss ℓ_A	Adversarial loss ℓ_G	PSNR	SSIM
Bicubic	–	–	–	24.91	0.667
S_1	✓	✓	✓	29.77	0.848
S_2	✓	✓		29.79	0.849
S_3	✓		✓	29.72	0.845
S_4		✓	✓	29.74	0.846
S_5	✓			29.74	0.848
S_6		✓		29.79	0.849

IV. EXPERIMENTS

A. IMPLEMENTATION DETAILS

In our LightGAN network, every convolution layer is composed of 64 filters with $3 \times 3 \times 3 \times 3$ dimension in both spatial and angular dimensions, respectively. The filters are initialized using the method of Glorot and Bengio [49]. Furthermore, we use the layout of the residual block proposed by Gross and Wilber [50] with two 4D convolutional layers, followed by batch normalization and the LReLU with a slope $\alpha = 0.2$ in the negative domain as the nonlinear activation function.

We randomly select 100 scenes from the dataset Lytro Archive [31] (excluding the Reflective and Occlusions categories) and Fraunhofer dataset [51] for training. The former has 353 light field images that capture real-world scenes with a Lytro Illum camera. In our experiment, we only select the center 9×9 views to avoid the dark SAIs at the corner. The latter is composed of 9 scenes captured with an ordinary camera moving along two orthogonal directions (angular dimension). The camera records the images from 21 vertical camera positions and 101 horizontal camera positions, which results in light fields with 21×101 angular resolution.

In our experiments, the low-resolution training data are obtained by downsampling according to

$$I^L = \delta\left(\kappa * I^H\right) + \eta, \tag{8}$$

where $\delta(\cdot)$ is the nearest neighbor downsampling operator applied on each view, κ denotes a Gaussian blurring kernel with a window size of 7×7 and standard deviation of 1.2 pixels. We also put in additive noise η, with zero mean and unit standard deviation.

B. TRAINING DETAILS

Our network takes a 4D patch of light field as the input and outputs the corresponding super-resolved 4D patch. The entire framework is implemented using the Tensorflow toolbox and trained with the Adam optimizer. Initially, the learning rate is set to 10^{-5} and reduced by a factor of 0.1 every 10 epochs. The generator network and discriminator network are updated alternately during the testing process until the batch-normalization layer.

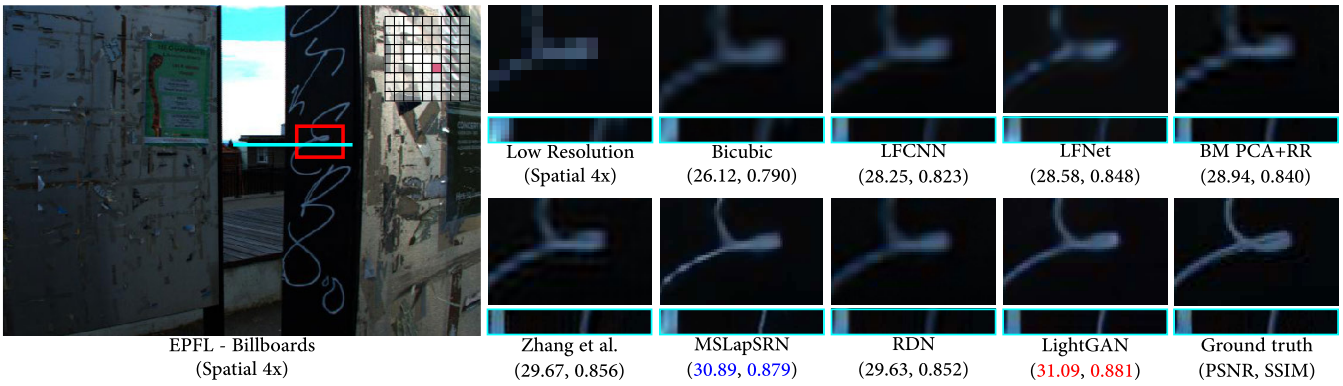


FIGURE 6. Visual comparison for spatial 4× SR task on the real-world scene. We report the average PSNR and SSIM value for each algorithm. The red text denotes the best result, while the blue one denotes the second best result.

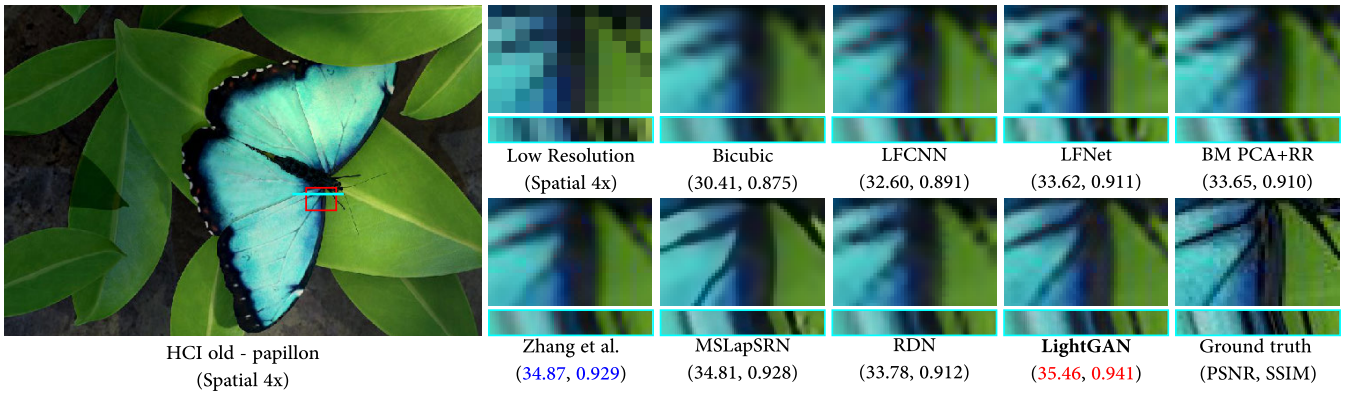


FIGURE 7. Visual comparison for spatial 4× SR task on the synthetic scene. We report the average PSNR and SSIM value for each algorithm. The red text denotes the best result, while the blue one denotes the second best result.

C. EFFECTIVENESS OF THE LightGAN

To evaluate the effectiveness of our model, we conduct the ablation studies on different experimental settings which are exhibited in Table 1. The proposed LightGAN is trained with all three loss terms (S_1), including the *spatial conceptual loss* ℓ_S, the *angular correlation loss* ℓ_A and the *generative adversarial loss* ℓ_G. The evaluations of our ablation studies are from two aspects. First, to figure out the contribution of GAN framework, we exam the performances of the generator trained with a discriminator or trained individually. In Table 1, the settings without adversarial loss (i.e. S_2, S_5 and S_6) mean that the generator is trained individually in the corresponding experiments. On the other hand, in the experiments with the settings S_1, S_3, and S_4, the generator is trained together with a discriminator. The Bicubic setting is used as the baseline. By comparing the quantitative results of settings S_1 and S_2, S_3 and S_5, S_4 and S_6, one can find that the quantitative values tend to drop when adding the adversarial loss for training. However, according the visual results presented in Fig. 8, such a loss term on the other hand contributes to more plausible visual results. Second, the ablation studies also evaluate the impacts of different loss terms on the network performance. For example, the angular correlation loss ℓ_A encourages the network to reconstruct high PSNR results which however, also tend to contain smooth artifacts in the texture regions. The spatial conceptual loss ℓ_S can also contribute to the visual results and one witnesses the improvements by comparing the columns S_2 and S_6 in Fig. 8.

D. RESULTS AND ANALYSES

The proposed LightGAN learns to approximate the original light field distribution, which allows it to deal with both spatial and angular low-resolution problems. Therefore, we conduct the evaluation on multiple tasks to illustrate the effectiveness of our generative method.

For spatial resolution enhancement, we evaluate our method in two aspects. First, we compare the performance against several advanced methods specially designed for light field, including LFCNN [19], LFNet [20], BM PCA+RR [37], and Zhang *et al.* [36]. LFCNN enhances the spatial and angular resolution separately, and for each SAI, the model only makes use of the parallax information from two neighboring views and omits the other views. In addition, they use a primitive framework for the spatial recovery, which results in a rough reconstruction. In both Fig. 6 and Fig. 7, their results are quite close to the baseline method (bicubic interpolation). Likewise, Wang *et al.* [20] attempt to iteratively model the two adjacent views

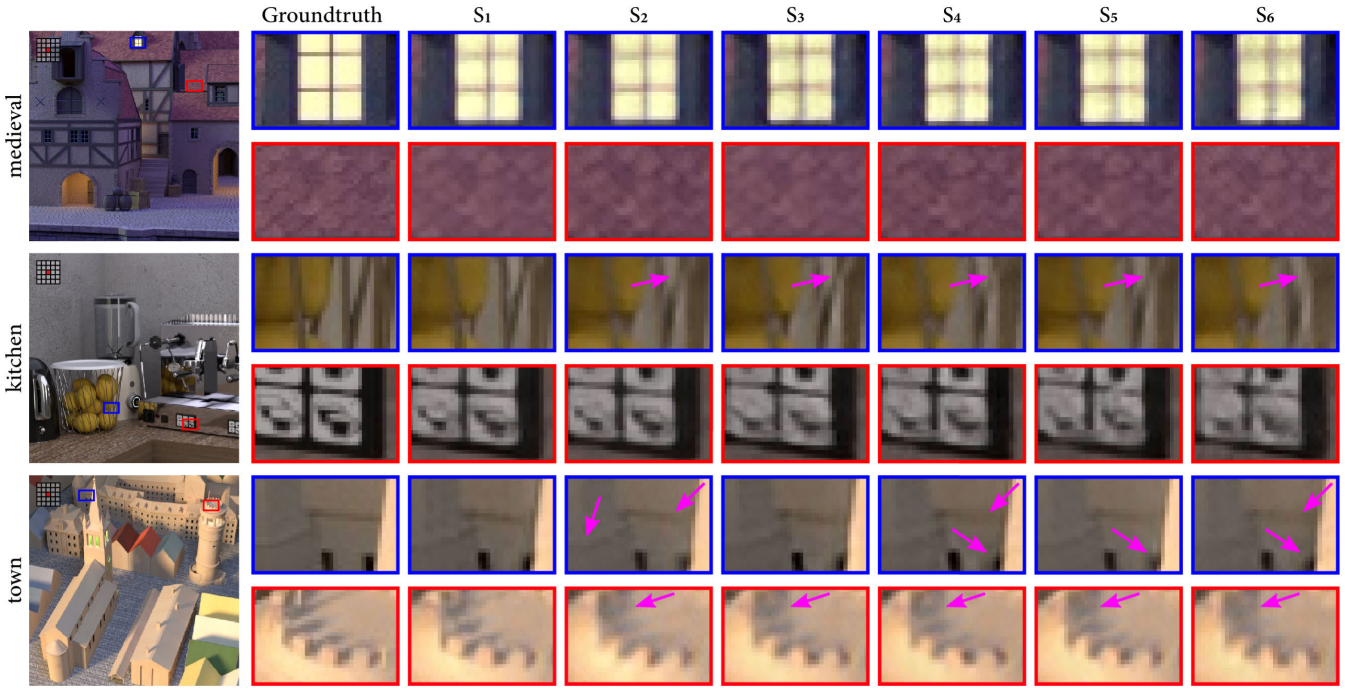


FIGURE 8. The visual reconstruction results correspond to the experiments with different settings in Table 1. As shown in the figure, the adversarial loss tends to drive the generator to produce more plausible results by comparing the columns S_1 and S_2, columns S_3 and S_5, columns S_4 and S_6.

horizontally and vertically with LFNet. Benefitting from a more sophisticated structure, their model achieves a slightly higher quantitative results than LFCNN.

There are also approaches considering more than two views for spatial reconstruction. Farrugia *et al.* [37] first rearrange the SAIs to obtain an image sequence, and then train their model to learn the linear projections between subspaces based on the patch volumes of the sequence. Unfortunately, the rearrangement has already destroyed the inherent structural property of the light field in the first step, let alone the limited representative capacity of the linear projection. On the other hand, Zhang *et al.* [36] choose to learn the recovery mapping (as described in Eq. 1) directly with a branched residual network. Each branch approximates the correlations of views in one direction (e.g. horizontal, vertical, and diagonal). The learned features are finally combined to predict the HR light field. All of these techniques exploit a pair or a sequence of SAIs each time, which more or less simplifies the complex correlations among the views.

By contrast, LightGAN can fully compute the entire spatio-angular information and therefore conduct an accurate reconstruction. This can be reflected both in spatial details recovery and EPI pattern recovery. Fig. 6 and Fig. 7 present the visual reconstruction results for spatial 4× task on real-world and synthetic scenes, respectively. Compared with other super-resolution methods, LightGAN can generate clear texture and maintain the epipolar property of the data. Quantitative comparison is presented in Table 2, which lists the peak signal-to-noise ratios (PSNR) values of different algorithms to illustrate their reconstruction results on both real-world and synthetic scenes.

TABLE 2. Quantitative evaluation (PSNR) on synthetic light field and real-world light field for $\gamma_s = 4$. All numbers are measured in dB. Not all of the Reflective and Occlusions scenes are used. In this study, we randomly select 20 scenes from each category for evaluation and report of the average PSNR values.

	Synthetic		Real-world	
	Buddha	Mona	Reflective (20)	Occlusions (20)
Bicubic	28.58	29.44	31.19	28.52
LFCNN [19]	29.84	31.40	31.42	28.86
LFNet [20]	30.93	32.47	33.85	30.37
BM PCA+RR [37]	30.43	32.68	33.07	30.45
Zhang et al. [36]	30.48	31.39	32.32	29.84
MSLapSRN [52]	30.98	32.74	32.43	30.85
RDN [53]	30.99	31.80	33.86	31.46
LightGAN	**31.93**	**32.97**	**35.24**	**33.15**

Given that the PSNR metric has its limitation to evaluate the fidelity of the results in terms of our visual sense, we make the second comparison with two state-of-the-art single image super-resolution methods, namely, MSLapSRN [52] and RDN [53]. These two algorithms are state-of-the-art methods particularly designed for single image super-resolution, and they are applied on each SAI individually to evaluate their performances on light field data. As shown in Fig. 6, the method provides more realistic details (the "doodles" on the billboards) compared with the other approaches. However, these two methods are designed for single image. That is they tend to omit the angular parallax information

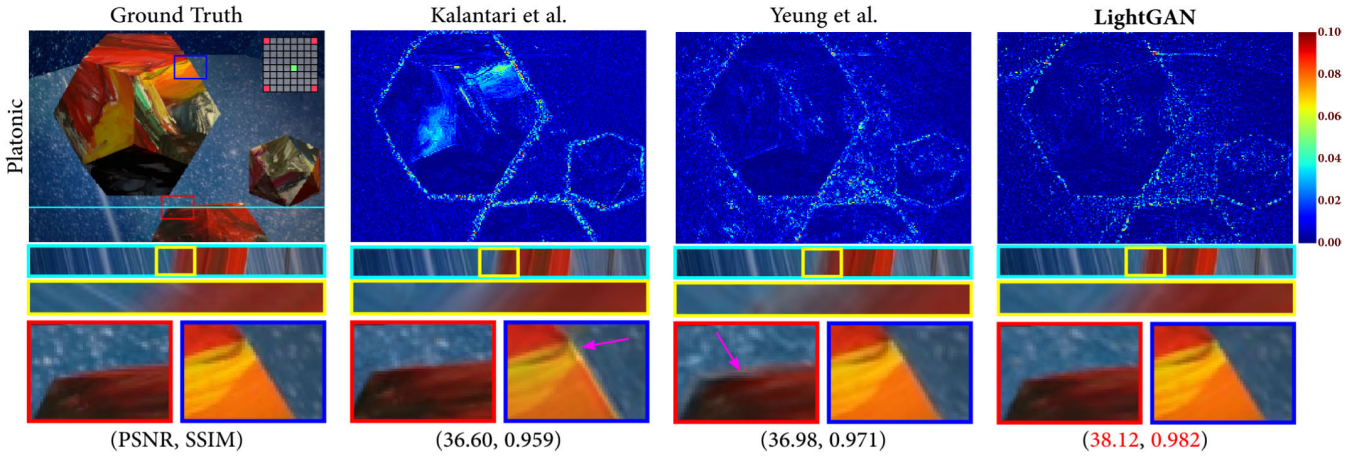


FIGURE 9. Visual results for 2 × 2 → 8 × 8 view synthesis task on a synthetic scene.

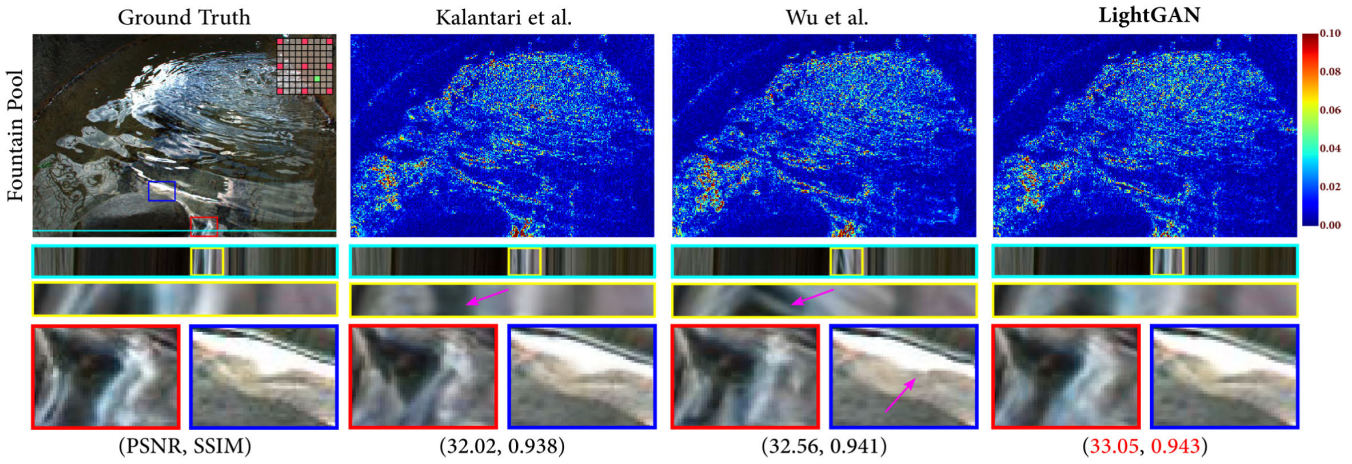


FIGURE 10. Visual results for 3 × 3 → 9 × 9 view synthesis task on a microscopy image.

which is proved to be helpful for spatial reconstruction by many multi-view algorithms. In contrast, our method is able to recover the details that quite approximate to the original ''doodles'' texture by considering both the spatial and angular information.

Besides spatial reconstruction, we also evaluate our method on view synthesis task. For each algorithm, we plot the residual map of the generated view to make the difference more visible. A darker blue means that the results resemble the ground truth more. In Fig. 9 and Fig. 10, we present the synthesis results on synthetic and real-world scenes, respectively. The first scene is from the HCI new dataset [54] (Platonic), which contains clear edges with a large range of depth. This scene is used to illustrate the performances of different algorithms on edge regions where occlusions appear. Compared with Kalantari *et al.* [35] and Yeung *et al.* [38], LightGAN can generate clearer edges as highlighted with arrows in Fig. 9.

The second scene is from the EPFL dataset (Fountain Pool), which contains a large region of non-Lambertian surface (water). Although the three comparison methods have competitive performances in terms of the residual maps, the LightGAN method provides more plausible spatial results (red box region in Fig. 10). Likewise, the EPI patterns in both figures are shown to demonstrate the recovery of correlations. Kalantari *et al.* [35] and Yeung *et al.* [38] tend to give rough spatial results, which subsequently impact their EPI results. Wu *et al.* [42] fail to recover the non-Lambertian surface, which leads to artifacts in the EPI pattern. In contrast, our GAN-based model produces more realistic spatial details with relatively sharp edges.

Fig. 9 and Fig. 10 also report the quantitative results. The presented PSNR and SSIM values are the average values calculated on each reconstructed views (exclude the input views denoted by red squares in both figures). As shown in the figures, although the proposed LightGAN is not trained to pursuit the high quantitative results, our model can still achieve the best performance compared with those learning models trained using the MSE loss [35], [38], [42]. In addition, we also test the runtime of our proposed method on a

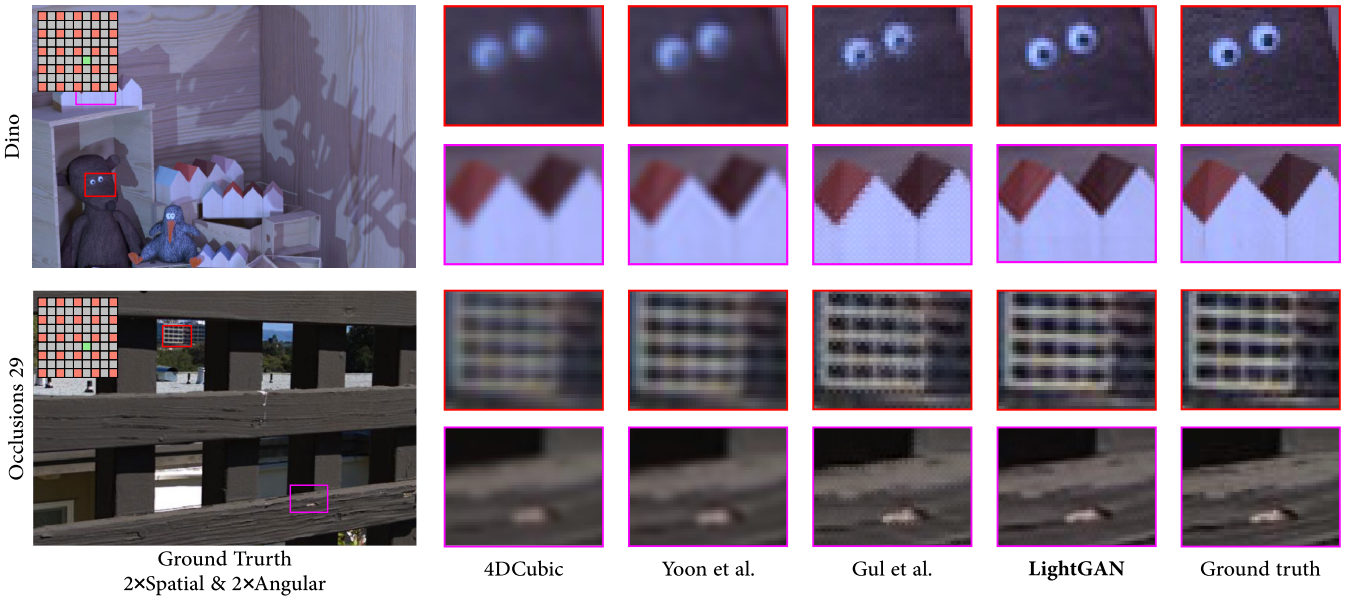


FIGURE 11. **Visual comparison for spatial-angular reconstruction. Both spatial and angular resolutions have been downsampled with the spatial factor $\gamma_s = 2$ and the angular factor $\gamma_a = 2$ (generate 9 × 9 views from 5 × 5 views).**

computer with and a NVIDIA Titan X GPU. The generator of LightGAN takes about 10 seconds to reconstruct of a light field with the resolution of 625 × 434 × 8 × 8 from a 625 × 434 × 2 × 2 light field input which is much faster than Kalantari *et al.* [35] and Wu *et al.* [42].

Last but not least, we evaluate the performance on spatio-angular recovery, i.e. 2× spatial and 2× angular enhancement. Method 4DCubic is presented as a baseline, and the label images are displayed in the last column in Fig. 11. As discussed in spatial comparison, LFCNN tends to provide a smooth reconstruction. Gul and-GunturkciteGul2018Spatial adopts multiple CNNs but they choose to conduct the recovery pixel-by-pixel. Such a strategy ignores the consistency of neighboring spatial pixels, and therefore results in the jagged artifacts near the object edges. Our algorithm, however, generates clearer edges, especially in the region with regular structures such as the "windows" and "roof" regions in Fig. 11.

V. CONCLUSIONS

In this paper, we propose a generative framework for light field reconstruction. In order to fully calculate the spatio-angular redundancy, we incorporate the HDC layers both in the generator and the discriminator, allowing the network to learn the direct mapping between the low-resolution and high-resolution light fields. By combining the angular correlation loss with the spatial perceptual loss and the adversarial loss, the proposed model is able to recover high-frequency spatial details with good visual fidelity.

The proposed generative framework is specially designed for light field processing and therefore it can handle a series of reconstruction problems, including spatial SR and view synthesis. We compare the performance of our model against state-of-the-art methods in both tasks. Experimental results show that our model is capable of simulating the local light field distribution in addition to enhancing the spatial or angular resolutions, and outperforms other state-of-the-art algorithms in most situations, especially when applied to scenes with complex occlusions and non-Lambertian surfaces.

REFERENCES

[1] Z. Xu, J. Ke, and E. Y. Lam, "High-resolution lightfield photography using two masks," *Opt. Express*, vol. 20, no. 10, pp. 10971–10983, May 2012.
[2] N. Chen, C. Zuo, E. Lam, and B. Lee, "3D imaging based on depth measurement technologies," *Sensors*, vol. 18, no. 11, p. 3711, Oct. 2018.
[3] K. Mitra and A. Veeraraghavan, "Light field denoising, light field superresolution and stereo camera based refocussing using a GMM light field patch prior," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2012, pp. 22–28.
[4] T.-C. Wang, A. A. Efros, and R. Ramamoorthi, "Depth estimation with occlusion modeling using light-field cameras," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 11, pp. 2170–2181, Nov. 2016.
[5] X. Sun, Z. Xu, N. Meng, E. Y. Lam, and H. K.-H. So, "Data-driven light field depth estimation using deep convolutional neural networks," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2016, pp. 367–374.
[6] P. P. Srinivasan, T. Wang, A. Sreelal, R. Ramamoorthi, and R. Ng, "Learning to synthesize a 4D RGBD light field from a single image," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2243–2251.
[7] N. Meng, T. Zeng, and E. Y. Lam, "Spatial and angular reconstruction of light field based on deep generative networks," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2019, pp. 4659–4663.
[8] C. Zhang, G. Hou, Z. Zhang, Z. Sun, and T. Tan, "Efficient auto-refocusing for light field camera," *Pattern Recognit.*, vol. 81, pp. 176–189, Sep. 2018.
[9] E. Y. Lam, "Computational photography with plenoptic camera and light field capture: Tutorial," *J. Opt. Soc. Amer. A, Opt. Image Sci.*, vol. 32, no. 11, pp. 2021–2032, Nov. 2015.
[10] T. E. Bishop, S. Zanetti, and P. Favaro, "Light field superresolution," in *Proc. IEEE Int. Conf. Comput. Photography (ICCP)*, Apr. 2009, pp. 1–9.
[11] J. Lim, H. Ok, B. Park, J. Kang, and S. Lee, "Improving the spatail resolution based on 4D light field data," in *Proc. 16th IEEE Int. Conf. Image Process. (ICIP)*, Nov. 2009, pp. 1173–1176.
[12] P. Didyk, P. Sitthi-Amorn, W. Freeman, F. Durand, and W. Matusik, "Joint view expansion and filtering for automultiscopic 3d displays," *ACM Trans. Graph.*, vol. 32, no. 6, p. 221, 2013.

[13] S. Vagharshakyan, R. Bregovic, and A. Gotchev, "Light field reconstruction using shearlet transform," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 1, pp. 133–147, Jan. 2018.
[14] N. Meng, E. Y. Lam, K. K. Tsia, and H. K.-H. So, "Large-scale multi-class image-based cell classification with deep learning," *IEEE J. Biomed. Health Informat.*, vol. 23, no. 5, pp. 2091–2098, Sep. 2019.
[15] Y. Yang, H. Chen, and J. Shao, "Triplet enhanced AutoEncoder: Model-free discriminative network embedding," in *Proc. 28th Int. Joint Conf. Artif. Intell.*, Aug. 2019, pp. 5363–5369.
[16] T.-C. Wang, J.-Y. Zhu, E. Hiroaki, M. Chandraker, A. A. Efros, and R. Ramamoorthi, "A 4D light-field dataset and CNN architectures for material recognition," in *Proc. Eur. Conf. Comput. Vis.*, Oct. 2016, pp. 121–138.
[17] N. Meng, X. Sun, H. K.-H. So, and E. Y. Lam, "Computational light field generation using deep nonparametric Bayesian learning," *IEEE Access*, vol. 7, pp. 24990–25000, 2019.
[18] N. Meng, H. K.-H. So, X. Sun, and E. Lam, "High-dimensional dense residual convolutional neural network for light field reconstruction," *IEEE Trans. Pattern Anal. Mach. Intell.*, early access, Oct. 1, 2019, doi: 10.1109/TPAMI.2019.2945027.
[19] Y. Yoon, H.-G. Jeon, D. Yoo, J.-Y. Lee, and I. S. Kweon, "Light-field image super-resolution using convolutional neural network," *IEEE Signal Process. Lett.*, vol. 24, no. 6, pp. 848–852, Jun. 2017.
[20] Y. Wang, F. Liu, K. Zhang, G. Hou, Z. Sun, and T. Tan, "LFNet: A novel bidirectional recurrent convolutional neural network for light-field image super-resolution," *IEEE Trans. Image Process.*, vol. 27, no. 9, pp. 4274–4286, Sep. 2018.
[21] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, "Photo-realistic single image super-resolution fusing a generative adversarial network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 105–114.
[22] M. Mathieu, C. Couprie, and Y. LeCun, "Deep multi-scale video prediction beyond mean square error," in *Proc. Int. Conf. Learn. Represent.*, 2016, pp. 1–17.
[23] S. Wanner and B. Goldluecke, "Variational light field analysis for disparity estimation and super-resolution," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 3, pp. 606–619, Mar. 2014.
[24] S. Pujades, F. Devernay, and B. Goldluecke, "Bayesian view synthesis and image-based rendering principles," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 3906–3913.
[25] T. E. Bishop and P. Favaro, "The light field camera: Extended depth of field, aliasing, and superresolution," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 5, pp. 972–986, May 2012.
[26] Y. Wang, G. Hou, Z. Sun, Z. Wang, and T. Tan, "A simple and robust super resolution method for light field images," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2016, pp. 1459–1463.
[27] M. Levoy and P. Hanrahan, "Light field rendering," in *ACM Conf. Comput. Graph. Interact. Techn.*, 1996, pp. 31–42.
[28] Z. Lin and H.-Y. Shum, "A geometric analysis of light field rendering," *Int. J. Comput. Vis.*, vol. 58, no. 2, pp. 121–138, Jul. 2004.
[29] C.-K. Liang and R. Ramamoorthi, "A light transport framework for lenslet light field cameras," *ACM Trans. Graph.*, vol. 34, no. 2, p. 16, Feb. 2015.
[30] D. Cho, M. Lee, S. Kim, and Y.-W. Tai, "Modeling the calibration pipeline of the lytro camera for high quality light-field image reconstruction," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 3280–3287.
[31] *Stanford Lytro Light Field Archive*. Accessed: Oct. 2018. [Online]. Available: http://lightfields.stanford.edu/
[32] M. Rerabek and T. Ebrahimi, "New light field image dataset," in *Proc. 8th Int. Conf. Qual. Multimedia Exper.*, Jun. 2016, pp. 1–7.
[33] G. Wu, Y. Liu, L. Fang, Q. Dai, and T. Chai, "Light field reconstruction using convolutional network on EPI and extended applications," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 7, pp. 1681–1694, Jul. 2018.
[34] N. Meng, X. Wu, J. Liu, and E. Lam, "High-order residual network for light field super-resolution," in *Association for the Advancement of Artificial Intelligence*. Palo Alto, CA, USA: AAAI Press, 2020.
[35] N. K. Kalantari, T.-C. Wang, and R. Ramamoorthi, "Learning-based view synthesis for light field cameras," *ACM Trans. Graph.*, vol. 35, no. 6, p. 193, 2016.
[36] S. Zhang, Y. Lin, and H. Sheng, "Residual networks for light field image super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 11046–11055.
[37] R. A. Farrugia, C. Galea, and C. Guillemot, "Super resolution of light field images using linear subspace projection of patch-volumes," *IEEE J. Sel. Topics Signal Process.*, vol. 11, no. 7, pp. 1058–1071, Oct. 2017.
[38] H. W. F. Yeung, J. Hou, J. Chen, Y. Y. Chung, and X. Chen, "Fast light field reconstruction with deep coarse-to-fine modeling of spatial-angular clues," in *Proc. Eur. Conf. Comput. Vis.*, Sep. 2018, pp. 137–152.
[39] Y. Wang, F. Liu, Z. Wang, G. Hou, Z. Sun, and T. Tan, "End-to-end view synthesis for light field imaging with pseudo 4DCNN," in *Proc. Eur. Conf. Comput. Vis.*, Sep. 2018, pp. 333–348.
[40] I. Goodfellow, "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst.*, Dec. 2014, pp. 2672–2680.
[41] G. Wu, B. Masia, A. Jarabo, Y. Zhang, L. Wang, Q. Dai, T. Chai, and Y. Liu, "Light field image processing: An overview," *IEEE J. Sel. Topics Signal Process.*, vol. 11, no. 7, pp. 926–954, Oct. 2017.
[42] G. Wu, M. Zhao, L. Wang, Q. Dai, T. Chai, and Y. Liu, "Light field reconstruction using deep convolutional network on EPI," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1638–1646.
[43] R. A. Farrugia and C. Guillemot, "Light field super-resolution using a low-rank prior and deep convolutional neural networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 5, pp. 1162–1175, May 2018.
[44] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
[45] A. L. Maas, A. Y. Hannun, and A. Y. Ng, "Rectifier nonlinearities improve neural network acoustic models," in *Proc. Int. Conf. Mach. Learn.*, vol. 30, no. 1, 2013, p. 3.
[46] W. Shi, J. Caballero, F. Huszar, J. Totz, A. P. Aitken, R. Bishop, D. Rueckert, and Z. Wang, "Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 1874–1883.
[47] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. Int. Conf. Learn. Represent.*, 2015, pp. 1–4.
[48] N. Meng, T. Zeng, and E. Y. Lam, "Perceptual loss for light field reconstruction in high-dimensional convolutional neural networks," in *Proc. Imag. Appl. Opt.*, 2019, p. 5.
[49] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proc. 13th Int. Conf. Artif. Intell. Statist.*, 2010, pp. 249–256.
[50] S. Gross and M. Wilber, "Training and investigating residual nets," *Facebook AI Res.*, vol. 6, May 2016.
[51] M. Ziegler, R. op het Veld, J. Keinert, and F. Zilly, "Acquisition system for dense lightfield of large scenes," in *Proc. 3DTV Conf., True Vis.*, Jun. 2017, pp. 1–4.
[52] W.-S. Lai, J.-B. Huang, N. Ahuja, and M.-H. Yang, "Fast and accurate image super-resolution with deep Laplacian pyramid networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 11, pp. 2599–2613, Nov. 2019.
[53] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu, "Residual dense network for image super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 2472–2481.
[54] K. Honauer, O. Johannsen, D. Kondermann, and B. Goldluecke, "A dataset and evaluation methodology for depth estimation on 4D light fields," in *Proc. Asian Conf. Comput. Vis.*, Nov. 2016, pp. 19–34.
[55] M. S. K. Gul and B. K. Gunturk, "Spatial and angular resolution enhancement of light fields using convolutional neural networks," *IEEE Trans. Image Process.*, vol. 27, no. 5, pp. 2146–2159, May 2018.

NAN MENG received the bachelor's degree from the University of Electronic Science and Technology of China, in 2015. He is currently pursuing the Ph.D. degree with the Department of Electrical and Electronic Engineering, University of Hong Kong. His research interests include machine learning, light field reconstruction, light field rendering, and medical imaging.

ZHOU GE received the B.S. degree in communication engineering from Fudan University, in 2014, and the M.S. degree in electrical engineering from the Imperial College London, in 2015. He is currently pursuing the Ph.D. degree with the Department of Electrical and Electronic Engineering, University of Hong Kong. His research interests include 3D computational imaging, machine learning, and image processing.

TIANJIAO ZENG received the B.S. degree in electrical engineering from the University of Electronic Science and Technology of China, and the M.S. degree in electrical and computer engineering from Rutgers University. She is currently pursuing the Ph.D. degree with the Department of Electrical and Electronic Engineering, University of Hong Kong. Her research interests include pattern recognition, computational imaging, and machine learning.

EDMUND Y. LAM (Fellow, IEEE) received the B.S., M.S., and Ph.D. degrees in electrical engineering from Stanford University. From 2010 to 2011, he was a Visiting Associate Professor with the Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology. He is currently a Professor in electrical and electronic engineering with the University of Hong Kong, and serves as the Computer Engineering Program Director. His main research is in computational imaging. He is also a Fellow of the OSA, SPIE, IS&T, and HKIE. He was a recipient of the IBM Faculty Award.