Pattern Recognition 171 (2026) 112147

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/pr

Disparity guidance and spatial-angular interaction for single-view-based light field synthesis

Yifan Yang [a,b], Zhen Qiu [a], Shuhai Zhang [a], Mingkui Tan [a,c,*]

[a] School of Software Engineering, South China University of Technology, Guangzhou, China
[b] Present address: Guangdong Provincial Key La-boratory of Power System Network Security, Electric Power Research Institute, CSG, Guangzhou, China
[c] Key Laboratory of Big Data and Intelligent Robot (South China University of Technology), Ministry of Education, Guangzhou, China

ARTICLE INFO

Keywords:
Light field
Generative adversarial network
Novel view synthesis

ABSTRACT

Local light field synthesis aims to synthesize a 4D light field from a single-view image, and it has a vast range of applications, such as VR, refocusing and depth estimation. This task, however, is very challenging due to the difficulty in understanding the scene geometry (e.g.disparity map) from a single-view image. Existing methods tackle this task either by warping the estimated scene geometry, or by super-resolution technique. Nevertheless, they may fail due to the inaccurate estimation of scene geometry without the supervision of ground truth and insufficient usage of angular and spatial information of light fields. To address these, we propose a Disparity-Guided Light Field generation (DGLF) paradigm. Specifically, DGLF parallelly generates target light fields and disparity maps which are exploited to guide the generation process to be aware of scene geometry. Moreover, we devise an interactive discriminator that aids in synthesizing a more perceptually realistic light field by capturing the interaction of spatial and angular information. Lastly, based on DGLF and interactive discriminator, we develop a disparity-guided adversarial generative network (DG-GAN) for light field synthesis. We theoretically analyze the generalization performance of DGLF. Extensive experiments on five light field datasets demonstrate the effectiveness of DG-GAN. Our code link: https://github.com/YifYang993/DG-GAN.git

1. Introduction

Light-field imaging has become a powerful technology that captures rich visual information about our world [1]. Recently, light fields have gained great attention with its exciting applications in Virtual Reality (VR) [2], refocusing [3], and depth estimation [4]. However, it is difficult for traditional cameras and cell phones to capture light fields, while commercial light field cameras are expensive and not widespread enough. To popularize the light field, we seek to synthesize a 4D light field from a single-view image, which is termed local light field synthesis (LLFS) [5]. LLFS is particularly challenging since it is non-trivial to understand the geometric structure of a scene from a single-view image.

Existing LLFS methods mainly fall into the following two categories. 1) Geometry-based approaches [5–8] first estimate the scene geometry (e.g. disparity map, point cloud, and appearance flow) and then use the geometry to generate target light fields. Specifically, Srinivasan et al. [5] and seo et al. [8] first estimate disparity maps of novel views and then warp target images based on the disparity maps. Li et al. introduce a pre-estimated disparity map as additional input and synthesize a target light field via composing multiplane images (MPI) under visible and occluded situations. 2) Super-resolution-based approaches [9,10] tackle LLFS via a single image super-resolution technique. For example, Chen et al. [9] super-resolve a given image to a light field by generative adversarial networks without explicitly estimating disparity maps.

There exist two potential issues in recent LLFS methods: 1) geometry-based methods [5–7] (See Fig. 1 (a)) may obtain distorted target light fields due to the inaccurate estimation of geometry in the absence of supervision [11]; 2) super-resolution-based methods [9,10] usually fail to further exploit both spatial and angular information of the light field, leading to inferior performance on light field synthesis. For instance, Wang et al. [12] have demonstrated that the complementarity among angular information ignored by super-resolution-based methods are able to enhance the extraction of spatial information. Given the first issue, a natural and critical question is raised. *Can the negative effects caused by inaccurate intermediate results (e.g., disparity maps) be mitigated when enabling the model to be aware of the geometry structure of light fields?* We attempt to answer this question by devising a Disparity-Guided Light Field generation (DGLF) paradigm. As shown in Fig. 1 (b), the core idea is to exploit estimated geometries to guide the generation of light fields,

* Corresponding author.
E-mail address: mingkuitan@scut.edu.cn (M. Tan).

https://doi.org/10.1016/j.patcog.2025.112147
Received 1 October 2024; Received in revised form 6 July 2025; Accepted 11 July 2025
Available online 20 July 2025

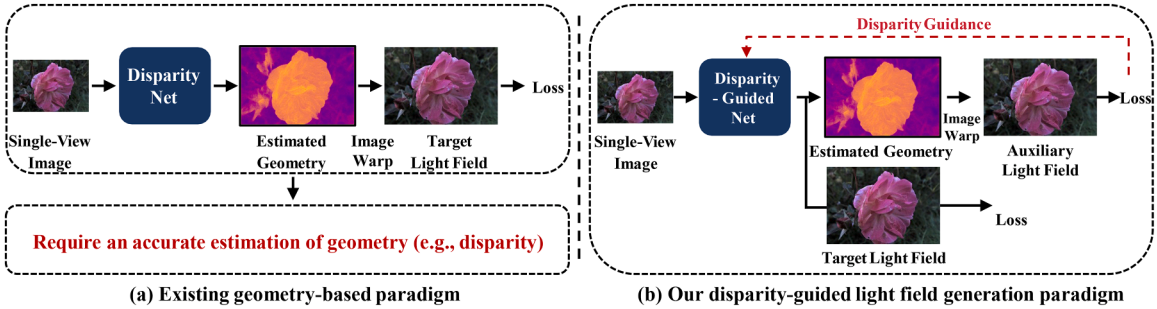


Fig. 1. Motivation for the disparity-guided paradigm. Existing geometry-based methods (a) depend on precise geometric structure estimation (e.g., disparity maps), which typically lack ground truth supervision. To address this, we propose a novel paradigm (b) that reduces reliance on accurate geometric estimation while preserving geometric awareness through disparity guidance. .

rather than explicitly using disparity maps to synthesize target light fields. Specifically, DGLF predicts the geometry and target light fields simultaneously by a shared model. Based on the generated geometry, an auxiliary light field is obtained via image warp. By minimizing the reconstruction errors of the auxiliary light field, the geometry information is implicitly embedded into the model and thus helps to synthesize target light fields more accurately. Since the estimated geometries are not explicitly used to synthesize the target light field, DGLF circumvents the estimation of accurate disparities to generate light fields.

To address the second issue, we seek to synthesize more perceptually-realistic light fields by exploring the interaction between spatial and angular information. To this end, we devise an interactive discriminator where we further introduce a spatial-angular interaction module to extract and incorporate the spatial and angular features of generated light fields. In this way, the discriminator is enhanced for judging the generated light fields from real light fields. As a result, more perceptually realistic light fields can be obtained. We empirically show that the interactive discriminator helps to generate a more photo-realistic light field.

Based on DGLF and the interactive discriminator, we propose a disparity-guided generative adversarial network (DG-GAN). Considering DG-GAN is not limited to specific network architecture, we apply it to two different neural networks, i.e. SRGAN [13] and DRN [14] for evaluation. We empirically show that DG-GAN achieves a more photo-realistic light field and fewer outliers when compared with state-of-the-art methods. Moreover, quantitative results on five different light field datasets demonstrate the effectiveness of DG-GAN. The main contributions of this work are threefold:

- We propose a Disparity-Guided Light Field generation (DGLF) paradigm for local light field synthesis. The paradigm facilitates the model to be aware of the geometry structure, and meanwhile mitigates the adverse effects of inaccurate estimation of geometry when synthesizing the target light field. We empirically demonstrate the superiority of DGLF in capturing scene geometry.
- We devise an interactive discriminator to enhance the perceptual reality of light fields by capturing the interaction of spatial-angular information.
- Based on the proposed DGLF and the interactive discriminator, we propose a disparity-guided generative adversarial network (DG-GAN) for local light field synthesis. Promising results on five datasets demonstrate DG-GAN is able to generate realistic and high-quality light fields.

2. Related work

To synthesize a 4D light field from multiple images or from a single-view image. some works introduce geometric information, e.g. disparity map [5,15–17] in the process of visual synthesis, while others [9,12] do not consider.

2.1. Light field synthesis with geometry structure

Among various geometric estimates, we mainly discuss the disparity estimation. Kalantari et al. [15] first propose to synthesize a light field by deep networks using two CNNs to estimate the disparity and color. Considering the generated disparity maps are unsupervised, following works [18] introduce implicit supervision by warping the estimated disparity map to synthesize light fields. To model the disocclusions and non-Lambertian effects, Li et al. [6] propose to synthesize light field via two multiplane image (MPI) networks to handle visible and occluded areas. With the flourish of transformer-based models, LF-DGT [17], a Transformer-based network that uses disparity-guided self-attention windows, relative positional encoding, and a gating mechanism to integrate light field structural priors for spatial super-resolution. To promote the light field in real-life applications, Srinivasan et al. [5] first propose the task of local light field synthesis (LLFS), which aims at using a single-view image to synthesize a 4D light field. Srinivasan et al. address LLFS by estimating disparities and generating the light field with image warp. We take inspiration from the way geometry-based methods learn geometry of a scene, but design a different way by only using the estimated geometry to guide the model for understanding the geometry. In this way, we mitigate the adverse effects of inaccurately estimated geometry.

2.2. Light field synthesis without geometry structure

The method of generating light fields without using geometric information is a major branch. Multi-view-based approaches usually leverage redundancy across views and use complementary information among views to learn the mapping from input to novel views. Yoon et al. [19] first apply CNNs to light field synthesis using complementary information among horizontal views and vertical views. To make fuller use of the information between each view, Gul et al. [20] propose to restore the spatial and angular resolution of a light field by CNNs. Although previous methods are effective in synthesizing novel views of light fields, they require multiple view images as input, which are usually inaccessible in real-world applications. Considering this, Chen et al. use a single-view image to synthesize a 4D light field by introducing a generative adversarial network [9]. However, they fail to further exploit both spatial and angular information of light fields, resulting in inferior perceptual quality regarding the synthesized light field. Here, we propose to synthesize a more photo-realistic light field by designing an interactive discriminator that captures the interaction of spatial and angular information.

2.3. Light field super-resolution

Recent advancements in light field super-resolution have seen methods exploiting internal information of light fields. For instance, Mo et al. [21] employ dual-attention networks for feature fusion. Li et al. [16] introduces a novel Epipolar Focus Spectrum (EFS) representation for

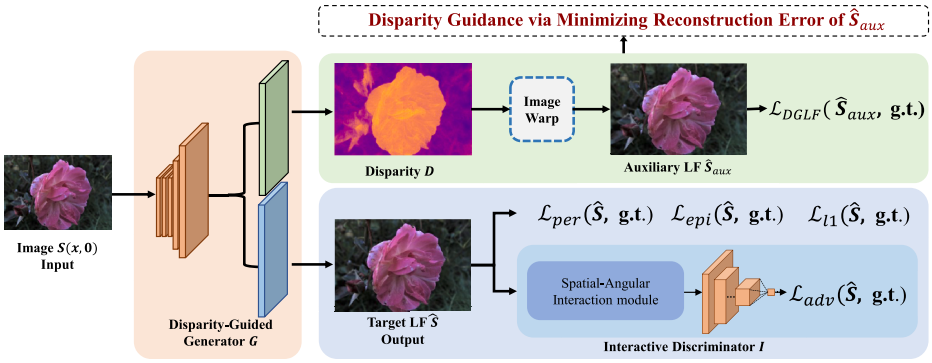


Fig. 2. Overview of our disparity-guided generative adversarial network. Our model simultaneously synthesizes the target light field and its corresponding disparity map. The disparity map is then used to generate an auxiliary light field via image warping. To enforce geometric consistency, we guide the model by minimizing the reconstruction loss of the auxiliary light field. Additionally, we introduce an interactive discriminator that captures the interplay between the spatial and angular information of the target light field.

light field reconstruction, effectively addressing occlusion and large disparity challenges while ensuring cross-view consistency. These works require multi-view images as input, while our method stands out in two aspects: 1) We synthesize light fields from a single-view image, reducing reliance on dense views. 2) Our approach enhances realism through an interactive discriminator that captures spatial-angular interactions.

3. Disparity-guided local light field synthesis

3.1. Problem definition of local light field synthesis.

Let $\mathcal{Z} = \{(\mathbf{S}(\mathbf{x}, \mathbf{0})_i, \mathbf{S}_i)\}_{i=1}^{N}$ be N training data with the center-view image $\mathbf{S}(\mathbf{x}, \mathbf{0}) \in S^{center}$ and the ground truth light field $\mathbf{S} \in S^{gt}$, where $\mathbf{x}$ is the spatial coordinate and $\mathbf{0}$ is the angular coordinate for the center view of the light field $\mathbf{S}$. Here, $\mathbf{S}(\mathbf{x}, \mathbf{0}) \in \mathbb{R}^{h \times w}$ and $\mathbf{S} \in \mathbb{R}^{u \times v \times h \times w}$, $u \times v$ is the angular dimension (e.g. $u = 4, v = 5$ for 4×5 light field) and $h \times w$ is the resolution of $\mathbf{S}(\mathbf{x}, \mathbf{0})$. The goal of Local Light Field Synthesis (LLFS) is to learn a function $\phi(\cdot)$ that maps the center-view image $\mathbf{S}(\mathbf{x}, \mathbf{0})$ to the target light field $\hat{\mathbf{S}}$ that is as similar to the ground truth light field $\mathbf{S}$ as possible:

$$\min_{\phi} ||\mathbf{S} - \hat{\mathbf{S}}||_1, \text{where } \hat{\mathbf{S}} = \phi(\mathbf{S}(\mathbf{x}, \mathbf{0})). \tag{1}$$

This task, however, is very challenging since it is difficult to understand the geometry structure of a scene from a single-view image. To tackle this, one feasible solution is to synthesize target light field with an estimated geometry (e.g. disparity map or point cloud). However, due to practical labeling difficulties, we may not have any annotations on the geometry. Thus, imprecise geometry estimation can introduce errors, leading to degradation in the rendered light fields [11]. In addition, another possible solution is to generate light fields in a super-resolution (SR) manner, which reduces the requirement on estimating an accurate geometry. However, SR-based methods tend to ignore the complementary information between the angular and spatial domains, and thus may lead to degraded performance.

3.2. Motivation and method overview

To address the above issues, we attempt to mitigate the adverse effects of inaccurate estimation of geometry when introducing geometry into a model. We hence first propose a novel paradigm, namely disparity-guided light field generation (DGLF), for learning the geometry structure of a light field from a single-view image. As shown in Fig. 2, we train a generator $G(\theta_G)$ to generate a target light field and a disparity map simultaneously. Then, we obtain an auxiliary light field by warping the disparity map. By minimizing the reconstruction error of the auxiliary light field, we introduce geometry information into the generator and alleviate the need for an accurate estimation of geometry (cf. Fig. 1 (b)). On the other hand, we attempt to generate a more photo-realistic light field by devising an interactive discriminator. The synthesized light field is fed into the interactive discriminator $I(\theta_I)$ for distinguishing from the ground truth light field by capturing the interaction of spatial and angular information. On the contrary, we train the disparity-guided generator $G(\theta_G)$ to output more accurate light fields in an adversarial manner. Based on DGLF and the interactive discriminator, we design a disparity-guided generative adversarial network (DG-GAN) for LLFS.

3.3. Optimization of local light field synthesis

To handle the task of local light field synthesis, We seek to train the proposed DG-GAN with the following loss functions: 1) an adversarial loss $\mathcal{L}_{adv}$ [13] for encouraging the generator to produce a more perceptually realistic light field; 2) a perceptual loss $\mathcal{L}_{per}$ [13] for generating the light field with perceptual satisfying solutions; 3) an epipolar image gradient loss $\mathcal{L}_{epi}$ for preserving the geometry consistency among each view of generated light fields; 4) a disparity-guided loss $\mathcal{L}_{DGLF}$ for guiding the model to be aware of geometry when synthesizing the target light field.

Adversarial loss. We impose the adversarial loss $\mathcal{L}_{adv}$ that enforces the synthesized light field to be consistent with the ground truth light field:

$$\mathcal{L}_{adv} = \mathbb{E}_{\mathbf{S} \sim P_{\mathbf{S}}}\left[\log I(\mathbf{S})\right] + \mathbb{E}_{\hat{\mathbf{S}} \sim p_{\hat{\mathbf{S}}}}\left[\log\left(1 - I(\hat{\mathbf{S}})\right)\right], \tag{2}$$

where $I(\hat{\mathbf{S}})$ is the probability that the discriminator I judges the synthesized light field $\hat{\mathbf{S}}$ to be a natural light field.

Perceptual loss. Following SRGAN [13], we use perceptual loss $\mathcal{L}_{per}$ to achieve perceptual realistic light field by comparing the perceptual features between generated and ground truth light fields:

$$\mathcal{L}_{per} = \frac{1}{W_{i,j} H_{i,j}} \sum_{x=1}^{W_{i,j}} \sum_{y=1}^{H_{i,j}} \left(\phi_{i,j}(\mathbf{S})_{x,y} - \phi_{i,j}(\hat{\mathbf{S}})_{x,y}\right)^2, \tag{3}$$

where $W_{i,j}$ and $H_{i,j}$ describe the dimensions of the respective feature maps $\phi_{i,j}(\mathbf{S})$ and $\phi_{i,j}(\hat{\mathbf{S}})$, and the features are extracted by the ImageNet-pretrained VGG-19 network.

Epipolar plane image (EPI) gradient loss. EPI [22] is a 2D slice of a 4D light field, and is constructed by fixing one dimension of spatial domain and angular domain. Here, we use EPI gradient loss $\mathcal{L}_{epi}$ [11] which

Table 1
Comparison of light field synthesis on five light field datasets. $DGGAN_S$ and $DGGAN_D$ denote our proposed DG-GAN using SRGAN [13] and DRN [14] as generator, respectively. Suffix "-Y" means our methods without using DGLF. "*" indicates methods using additional depth information from pretrained models.

Methods	General [15]		Flower [5]		HCI [25]		Stanford [26]		Inria [27]	
	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑
Srinivasan [5]	22.80	0.6303	25.01	0.6788	17.60	0.2850	10.94	0.3116	20.66	0.5512
Li et al. [6]*	18.88	0.7198	23.71	0.8108	9.97	0.1314	14.55	0.5354	20.69	0.6151
LFGAN [9]	21.65	0.6212	25.80	0.7498	19.54	0.2800	12.23	0.2100	23.66	0.6422
GenWarp [8]*	20.24	0.4964	21.86	0.4831	16.94	0.2870	9.37	0.4542	18.01	0.5517
ViewCrafter [28]*	18.58	0.4779	21.34	0.5326	15.42	0.2843	13.10	0.3692	14.00	0.3903
$DGGAN_{SY}$	23.27	0.7134	27.99	0.8179	20.77	0.3845	13.30	0.1609	23.89	0.6633
$DGGAN_S$ (Ours)	**23.50**	**0.7139**	**28.25**	**0.8351**	**21.23**	**0.3949**	**14.06**	**0.2337**	**24.18**	**0.6870**
$DGGAN_{DY}$	22.82	0.6799	28.97	0.8648	21.23	0.4393	13.96	0.5343	23.98	0.6710
$DGGAN_D$ (Ours)	**23.31**	**0.6800**	**29.11**	**0.8662**	**21.60**	**0.4489**	**14.03**	**0.5402**	**24.15**	**0.6975**

minimizes the L_1 distance between the gradient of EPIs from the synthesized light field and the ground truth light field:

$$\begin{aligned}\mathcal{L}_{epi} = &\sum_{y,v}\Big(\Big|\nabla_x \mathbf{E}_{y,v}(x,u) - \nabla_x \widehat{\mathbf{E}}_{y,v}(x,u)\Big| \\ &+ \Big|\nabla_u \mathbf{E}_{y,v}(x,u) - \nabla_u \widehat{\mathbf{E}}_{y,v}(x,u)\Big|\Big) \\ &+ \sum_{x,u}\Big(\Big|\nabla_y \mathbf{E}_{x,u}(y,v) - \nabla_y \widehat{\mathbf{E}}_{x,u}(y,v)\Big| \\ &+ \Big|\nabla_v \mathbf{E}_{x,u}(y,v) - \nabla_v \widehat{\mathbf{E}}_{x,u}(y,v)\Big|\Big),\end{aligned} \tag{4}$$

where $\mathbf{E}_{y,v}(x,u)$ and $\mathbf{E}_{x,u}(y,v)$ denote the ground truth horizontal and vertical EPI, respectively. $\widehat{\mathbf{E}}_{y,v}$ denotes the estimated horizontal EPI and ∇_v represents the image gradient w.r.t.dimension v.

To guide the model to be aware of geometry when synthesizing target light fields, in this paper, we propose a disparity-guided loss called $\mathcal{L}_{DGLF}$, which plays very important role in our method. For convenience, we leave the details of $\mathcal{L}_{DGLF}$ in the following subsections. The overall optimization problem for our Disparity-Guided Light Field Synthesis is formulated as follows:

$$\min_{\theta_G}\max_{\theta_I}\ \lambda_1\mathcal{L}_{adv}(\theta_G,\theta_I) + \lambda_2\mathcal{L}_{per}(\theta_G) + \lambda_3\mathcal{L}_{epi}(\theta_G) + \mathcal{L}_{DGLF}(\theta_G), \tag{5}$$

where θ_G and θ_I are parameters of the generator G and the discriminator I, respectively. Moreover, λ_1, λ_2 and λ_3 are trade-off parameters. To train the proposed model, we follow the training scheme of SRGAN [13] to optimize the above minimax objective.

3.4. Disparity-guided light field generation

To better acquire the geometric structure for guiding the single-view light field generation process, we design the disparity-guided light field generation paradigm to regulate the generation process by updating a geometry, i.e. disparity map $\mathbf{D}$. In this way, the synthesis of a light field is guided but not directly affected by an estimated disparity map. Without loss of generality, we describe the overall pipeline in a random 2D angular coordinate $\mathbf{u} = [u, v]$ of a light field $\mathcal{S}$, the remaining angular coordinate can be easily applied in same manner.

Given a center view image $\mathbf{S}(\mathbf{x}, \mathbf{0})$, a generator G consists of a shared feature mapping f and two groups of convolutional kernels g_S and g_D inside a 2D convolution. The generation of target light field $\hat{\mathbf{S}}$ and counterpart disparity $\mathbf{D}$ can be formulated as:

$$\hat{\mathbf{S}} = g_S(f(\mathbf{S}(\mathbf{x},\mathbf{0})),\ \mathbf{D} = g_D(f(\mathbf{S}(\mathbf{x},\mathbf{0})), \tag{6}$$

where $\mathbf{x}$ is the spatial coordinate (x, y), and $\mathbf{0}$ is the angular coordinate for the center view. In this way, the $\hat{\mathbf{S}}$ and $\mathbf{D}$ are separated in the final stage of generation.

Moreover, to guide the synthesis process with the estimated geometry, we warp $\mathbf{D}$ and $\mathbf{S}(\mathbf{x}, \mathbf{0})$, supervised by the ground truth light field $\mathbf{S}$, to generate an auxiliary light field $\hat{\mathbf{S}}_{aux}$:

$$\hat{\mathbf{S}}_{aux}(\mathbf{x},\mathbf{u}) = \mathbf{S}(\mathbf{x} + \mathbf{u}\mathbf{D}(\mathbf{x},\mathbf{u}),\mathbf{0}) = \mathrm{warp}(\mathbf{D}(\mathbf{x},\mathbf{u}),\mathbf{S}(\mathbf{x},\mathbf{0})), \tag{7}$$

where the warp operation denotes generating novel images by matching corresponding positions with disparity maps $\mathbf{D}$ in coordination $(\mathbf{x}, \mathbf{u})$. Here, we propose to implicitly introduce the geometry information into the shared feature mapping f by minimizing the reconstruction error of $\hat{\mathbf{S}}_{aux}$. The training process of DGLF aims to minimize two loss functions including a reconstruction loss $\mathcal{L}_{rec}$ for the target light field generation and a disparity-guided loss $\mathcal{L}_{guide}$ for assisting DG-GAN in understanding the geometry structure at angular coordinate $\mathbf{u}$ and spatial coordinate $\mathbf{x}$:

$$\mathcal{L}_{DGLF}(\mathbf{x},\mathbf{u}) = \underbrace{||\hat{\mathbf{S}}(\mathbf{x},\mathbf{u}) - \mathbf{S}(\mathbf{x},\mathbf{u})||_1}_{\text{reconstruction loss } \mathcal{L}_{rec}} + \lambda_0 \underbrace{||\mathbf{S}(\mathbf{x},\mathbf{u}) - \hat{\mathbf{S}}_{aux}(\mathbf{x},\mathbf{u})||_1}_{\text{disparity-guided loss } \mathcal{L}_{guide}}, \tag{8}$$

where λ_0 is a hyper-parameter for balancing $\mathcal{L}_{guide}$ and $\mathcal{L}_{rec}$. See the impact of λ_0 in Section 4.8.

Next, we analyze the generalization bound for our proposed disparity-guided light field generation paradigm (DGLF). To this end, we use the generalization error of DGLF with $\mathcal{L}_{DGLF}(\mathbf{x},\mathbf{u})$ to measure the accuracy of DGLF in predicting unseen light field data. Note that the loss function $\mathcal{L}_{DGLF}$ in Eqn. (8) is determined by f, g_S and g_D in Eqn. (6). Let $E(f, g_S, g_D) = \mathbb{E}[\mathcal{L}_{rec}(\mathbf{x},\mathbf{u}) + \lambda_0\mathcal{L}_{guide}(\mathbf{x},\mathbf{u})]$ and $\hat{E}(f, g_S, g_D)$ be its empirical loss. we obtain the generalization bound of the DGLF using Rademacher complexity [23].

Theorem 1 (Generalization Performance of DGLF). *Let $\mathcal{L}_{rec}(\mathbf{x},\mathbf{u}) + \lambda_0\mathcal{L}_{guide}(\mathbf{x},\mathbf{u})$ be a mapping from $\mathcal{S}^{center}\times\mathcal{S}^{gt}$ to $[0, B]$ with upper bound B and function space $\mathcal{H}^{DGLF}$. For any δ>0, all $(f, g_S, g_D)\in\mathcal{H}^{DGLF}$, with probability at least $1-\delta$, the generalization error $E(f, g_S, g_D)$, i.e., the expected loss, satisfies*

$$E(f,g_S,g_D) \le \hat{E}(f,g_S,g_D) + 2\mathrm{R}_N(\mathcal{H}^{DGLF}) + B\sqrt{\frac{1}{2N}\log\left(\frac{1}{\delta}\right)}, \tag{9}$$

where $\mathrm{R}_N(\mathcal{H}^{DGLF})$ is the Rademacher complexity of our proposed disparity-guided light field generator. Let $\mathcal{B}(f, g_S, g_D)$ be the generalization bound w.r.t. $\mathrm{R}_N(\mathcal{H}^{DGLF})$, i.e., $\mathcal{B}(f, g_S, g_D) = 2\mathrm{R}_N(\mathcal{H}^{DGLF}) + B\sqrt{\frac{1}{2N}\log\left(\frac{1}{\delta}\right)}$, then we have

$$\mathcal{B}(f,g_S,g_D) \le \mathcal{B}(f,g_S), \tag{10}$$

where $(f, g_S) \in \mathcal{H}^{LF}$, $\mathcal{B}(f, g_S)$ is the generalization bound without the DGLF paradigm w.r.t. the Rademacher complexity $\mathrm{R}_N(\mathcal{H}^{LF})$.

This theorem shows that the generalization bound of our proposed disparity-guided light field generation paradigm is smaller than the paradigm without it, which thus helps to synthesize more accurate light fields. We empirically validate this in Section 4.1 and demonstrate the results in Tables 1 and 2. Based on DGLF and the theoretical analysis, we naturally design a disparity-guided generative adversarial network to synthesize light fields.

Table 2
The impact of our disparity-guided light field generation paradigm and Inter-discriminator on light field synthesis on the General dataset. DGLF represents the guidance of disparity maps.

DGLF	w/o Discriminator			Traditional Discriminator [34]			Inter-Discriminator		
	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
✗	23.38	**0.7189**	0.3567	23.21	0.7036	0.3541	23.31	0.6926	0.3389
✓	**23.48**	0.7003	**0.3170**	**23.42**	**0.7109**	**0.3407**	**23.50**	**0.7139**	**0.3089**

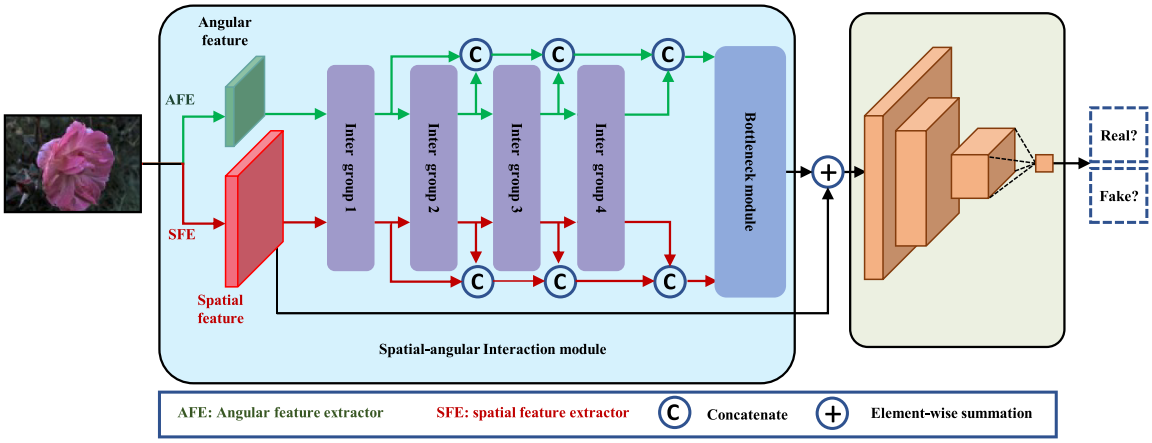


Fig. 3. An overview of our proposed interactive discriminator. Note that we introduce spatial-angular interaction module to our Interactive Discriminator for capturing and incorporating discriminative features in angular domain and spatial domain of generated light field.

3.5. Disparity-guided generative adversarial network

Driven by the DGLF, to suppress adverse effects of inaccurate geometry estimation while enabling the model to understand the geometry structure of light fields, we devise a disparity-guided generator for generating target light fields and disparity maps simultaneously. Moreover, to promote the vision reality of synthesized light fields, we design an interactive discriminator by exploiting the interaction between angular and spatial information of light fields. Combining the generator and the discriminator, we depict Disparity-Guided Generative Adversarial Network (DG-GAN) in detail.

Disparity-guided generator. Based on the disparity-guided light field generation paradigm, we aim to design a disparity-guided generator for the disparity map $\mathbf{D}$ and target light field $\hat{\mathbf{S}}$. Specifically, as shown in Fig. 2, a center view image $\mathbf{S}(\mathbf{x}, \mathbf{0})$ is first fed into a feature mapping f for feature learning and upsampling, resulting in a shared feature map $\mathbf{P}^{\mathbf{sc}}$. Note that to generate of $\mathbf{D}$ and $\hat{\mathbf{S}}$, we set the output channel of final layer convolution as two, so that two groups of convolutional kernels g_S and g_D inside the convolution are responsible for the generation of $\mathbf{D}$ and $\hat{\mathbf{S}}$, respectively. The learned feature map $\mathbf{P}^{\mathbf{sc}}$ is further mapped to $\mathbf{D}$ and $\hat{\mathbf{S}}$, which can be obtained by Eq . 6. Moreover, due to the flexibility of DG-GAN, disparity-guided generator can be implemented from existing super-resolution-based generator, e.g. SRGAN [13] and DRN [14] by adding one additional output channel on the last convolutional layer.

Interactive discriminator. Despite the successful generation of light fields by SR-based synthesis methods, they usually fail to further exploit both spatial and angular information of light fields [12], leading to inferior perceptual results. To encourage more realistic generation, one can introduce a standard discriminator to regularize the generator. However, such a scheme overlooks the underlying interactions between spatial and angular information. To address this, we devise a spatial-angular interactive discriminator (see Fig. 3) to enhance the generator by encoding the spatial-angular characteristic of light field, instead of directly discriminating light fields. Such a discriminator contains a spatial-angular interaction module, proposed in Wang et al. [12], for capturing the interaction between angular and spatial information, and groups of convolutional layers for feature learning and classification.

Suppose a synthesized macro-pixel image representation of light field $\hat{\mathbf{S}} \in \mathbb{R}^{b\times c\times hu\times wv}$, our goal is to fool the interactive discriminator, the discriminative ability of which is enhanced by capturing the spatial-angular interaction of light field with $\hat{\mathbf{S}}$.

We formulate the process of the interactive discriminator as the following equation:

$$\mathbf{p}_1^{\mathbf{a}} = \mathrm{AFE}(\hat{\mathbf{S}}),\ \mathbf{p}_1^{\mathbf{s}} = \mathrm{SFE}(\hat{\mathbf{S}}), \tag{11}$$

$$\mathbf{p}_4^{\mathbf{a}}, \mathbf{p}_4^{\mathbf{s}} = \mathrm{SAI}(\mathbf{p}_1^{\mathbf{a}}, \mathbf{p}_1^{\mathbf{s}}),\ \mathbf{p}^{\mathbf{as}} = \mathrm{Bot}(\mathbf{p}_4^{\mathbf{a}}, \mathbf{p}_4^{\mathbf{s}}), \tag{12}$$

$$\mathbf{p}^{\mathbf{c}} = \mathbf{p}^{\mathbf{as}} + \mathbf{p}_1^{\mathbf{s}},\ \mathrm{p} = \mathrm{C}(\mathbf{p}^{\mathbf{c}}). \tag{13}$$

From Fig. 3, the angular feature extractor AFE and spatial feature extractor SFE extract spatial feature $\mathbf{p}_1^{\mathbf{s}} \in \mathbb{R}^{\mathbf{b}\times\mathbf{c}_1\times\mathbf{hu}\times\mathbf{wv}}$ and angular feature $\mathbf{p}_1^{\mathbf{a}} \in \mathbb{R}^{\mathbf{b}\times\mathbf{c}_1\times\mathbf{h}\times\mathbf{w}}$, respectively. The spatial angular interaction module SAI aims to promote the information exchange between angular and spatial information. The bottleneck module Bot fuses the angular feature $\mathbf{p}_4^{\mathbf{a}}$ and spatial feature $\mathbf{p}_4^{\mathbf{s}}$, and C is the classifier to output the final probability p based on the fused feature $\mathbf{p}^{\mathbf{c}}$.

3.6. Detailed network architecture of DG-GAN

We implement $\mathrm{DGGAN_S}$, $\mathrm{DGGAN_{SY}}$ and $\mathrm{DGGAN_{SRGB}}$ with SRGAN [13]. The architectures of $\mathrm{DGGAN_D}$, $\mathrm{DGGAN_{DY}}$ and $\mathrm{DGGAN_{SRGB}}$ are implemented based on DRN [14], as shown in Table 9. We use the 8× version of DRN and SRGAN for synthesizing a light field with an angular resolution of 8×8. In order to make the details clearer and more concise, We only state the difference between our proposed network with SRGAN and DRN.

Architecture of $\mathrm{DGGAN_S}$ *and* $\mathrm{DGGAN_D}$. $\mathrm{DGGAN_S}$ and $\mathrm{DGGAN_D}$ are implemented based on SRGAN and DRN, respectively. The number of input channel in the first conv layer is set to 1 for super-resolving the Y

channel of a light field. To synthesize a disparity map and a light field simultaneously, we add one additional output channel on the last conv layer.

Architecture of DGGAN_{SY} *and* DGGAN_{DY}. To evaluate the effectiveness of our proposed DGLF paradigm. We design two variants, e.g. DGGAN_{SY}, DGGAN_{DY}, of DGGAN_{S} and DGGAN_{D}, respectively. The number of input channels for the first convolutional layer and output channels for the last convolutional layer in DGGAN_{SY} and DGGAN_{DY} are set to 1, for encoding and generating the target light field only.

Architecture of $\text{DGGAN}_{\text{SRGB}}$. For $\text{DGGAN}_{\text{SRGB}}$, since we compare it with existing single-view view synthesis (SVVS) methods, we follow the SVVS to handle light fields in RGB color space. To this end, we set the number of output channel for the last convolutional layer to 4, in which 3 serves the generation of RGB channel of a target light fields and 1 for a disparity map.

3.7. How the geometry information is embedded into our model

Our method adopts a Disparity-guided generation framework, consisting of a shared encoder f_θ with parameters θ and two branches: a generation branch g_{ϕ_1} with parameters ϕ_1 and an auxiliary branch g_{ϕ_2} with parameters ϕ_2. The generation branch g_{ϕ_1} synthesizes the target light field $\hat{\mathbf{S}}$, while the auxiliary branch g_{ϕ_2} estimates a geometry representation, i.e., a disparity map, and uses it to generate auxiliary light field $\hat{\mathbf{S}}_{aux}$ via an image warping operation. By explicitly estimating the disparity map, the auxiliary branch incorporates geometric awareness into the framework. Next, we analyze how geometric information is embedded into the target light field synthesis from a gradient update perspective.

Gradient flow in our framework. Given a single-view image $\mathbf{S}(\mathbf{x}, \mathbf{0})$, the shared encoder extracts a feature representation $h = f_\theta(x)$, which is then processed by the two branches to produce $\hat{\mathbf{S}} = g_{\phi_1}(h)$ and $\hat{\mathbf{S}}_{aux} = g_{\phi_2}(h)$. During training, both $\hat{\mathbf{S}}$ and $\hat{\mathbf{S}}_{aux}$ are supervised using the ground-truth light field $\mathbf{S}$. To enhance geometric reasoning, we introduce a guiding loss $\mathcal{L}_{guide} = \ell_1(\hat{\mathbf{S}}_{aux}, \mathbf{S})$ on the auxiliary branch, enforcing alignment with the ground-truth light field. The gradient update follows the chain rule:

$$\frac{\partial \mathcal{L}_{guide}}{\partial \theta} = \frac{\partial \mathcal{L}_{guide}}{\partial \hat{\mathbf{S}}_{aux}} \cdot \frac{\partial \hat{\mathbf{S}}_{aux}}{\partial h} \cdot \frac{\partial h}{\partial \theta}. \tag{14}$$

This update modifies θ, refining the shared feature representation h. Since the generation branch also relies on h, this update indirectly influences g_{ϕ_1} and improves the target light field synthesis:

$$\Delta \hat{\mathbf{S}} = \frac{\partial \hat{\mathbf{S}}}{\partial h} \cdot \frac{\partial h}{\partial \theta} \cdot \Delta\theta. \tag{15}$$

Thus, even though $\mathcal{L}_{guide}$ is not directly applied to g_{ϕ_1}, it enhances the representation learned by the shared encoder, leading to more accurate target light field synthesis. This interplay between branches highlights how our disparity-guided approach effectively integrates geometric information during inference.

3.8. Difference of our DG-GAN with existing methods

1) **Our method does not rely on accurate geometry**: Existing geometry-based method depends on precise geometry estimation for image warping and are often inaccurate in complex or occluded scenes. To alleviate the side-effect of inaccurate geometry, our approach simultaneously synthesizes target light fields and disparity maps, using the disparity to create an auxiliary light field through image warping. Note that we propose a paradigm to guide our generation model to be aware of geometry structure by minimizing reconstruction loss of an auxiliary light field (see Section 3.4).
2) **We further explore the spatial angular interaction of synthesized light field.** Existing super-resolution based methods usually fail to further exploit both spatial and angular information of the light field, leading to inferior performance on light field synthesis. Specifically, LFGAN treats light fields as a regular RGB image and sends them to a vanilla discriminator. Differently, we devise a spatial-angular interactive discriminator (see Section 3.5) to enhance the generator by encoding the spatial-angular characteristic of light fields, instead of directly discriminating light fields.

4. Experiments

Implementation details. We implement our approach in PyTorch[1] and train our networks using Adam with $\beta_1 = 0.9$ and $\beta_2 = 0.999$. The training are performed 120 epochs with batch size set to 1. The learning rates of the generator and the discriminator are set to 10^{-4}. and decayed at 60th epoch. The training and testing phases are performed on Nvidia GTX TITAN X GPU with 12 GB of memory. In all the experiments, following SRGAN [13], we set the λ_1 and λ_2 as 10^{-3} and 2×10^{-6}, respectively. Following [11], we set λ_3 to 1. For data processing, we perform light field synthesis by upscaling a single-view image $\mathbf{S}(\mathbf{x}, \mathbf{0}) \in \mathbb{R}^{h,w}$ with $8\times$ factor to generate target light fields $\hat{\mathbf{S}} \in \mathbb{R}^{8h,8w}$. we also follow [5,9] to use center 8×8 grids of a light field. In order to reduce the I/O burden when reading large size light field images, we use 128×128 patch of an input image for training and do not use any data augmentation strategy in all experiments.

Metrics for evaluation. The evaluation metrics PSNR, SSIM, and LPIPS are chosen for their complementary strengths in assessing light field synthesis. PSNR measures pixel-level accuracy, providing a straightforward quantitative measure of fidelity. SSIM evaluates structural similarity, aligning better with human perception by considering luminance, contrast, and structure. LPIPS, a learned perceptual metric, captures high-level semantic and fine-grained details using deep feature representations, closely matching human visual judgment. Together, these metrics ensure a balanced and robust evaluation, covering pixel accuracy, structural integrity, and perceptual quality, which are critical for comprehensive performance analysis.

Light field synthesis in YCbCr color space. Following [24], we conduct local light field synthesis in YCbCr color space using DGGAN_{D}, DGGAN_{S}, DGGAN_{DY} and DGGAN_{SY}, and only super-resolve Y channel of a center-view image. Specifically, $\mathbf{S}(\mathbf{x}, \mathbf{0})^{\text{Y}}$ is fed to DG-GAN to synthesize Y channel of target light field $\hat{\mathbf{S}}^{\text{Y}}$ and disparity map $\mathbf{D}$. Then, a CbCr channel image $\mathbf{S}^{\text{CbCr}}$ up-scaled from $\mathbf{S}(\mathbf{x}, \mathbf{0})^{\text{CbCr}}$ by bicubic interpolation is concatenated with $\hat{\mathbf{S}}^{\text{Y}}$ to obtain the target light field $\hat{\mathbf{S}}^{\text{YCbCr}}$. Here, we propose to generate novel views of the light field on y channel and directly upsample the remaining Cb Cr channel with bicubic interpolation.

Datasets. We evaluate our method on the following five datasets. 1) **Flower** [5] is a challenging dataset with complex occlusion. Such a dataset contains 3343 samples of various categories including roses, poppies, orchids, iris and other plants. Following Srinivasan et al. [5], we randomly divide the dataset into 3243 for training and 100 for testing. 2) **General** dataset [15] contains 102 samples of a variety of object categories such as seahorse, cars, stones in arbitrary location, which poses a challenge for the generalization of the method. We use the same spatial resolution and angular resolution as the flower dataset and following Kalantari et al. [15] to split the dataset into 72 for training and 30 for testing. 3) **HCI** [25] contrains 24 light field. HCI faces five challenges including occlusion boundary, fine structures, low texture, smooth surface and camera noise. We use 16 light fields for training and 8 for testing. 4) **Stanford** [26] consists of 9 samples with different spatial

[1] We recommend checking the generated light fields in supplementary.

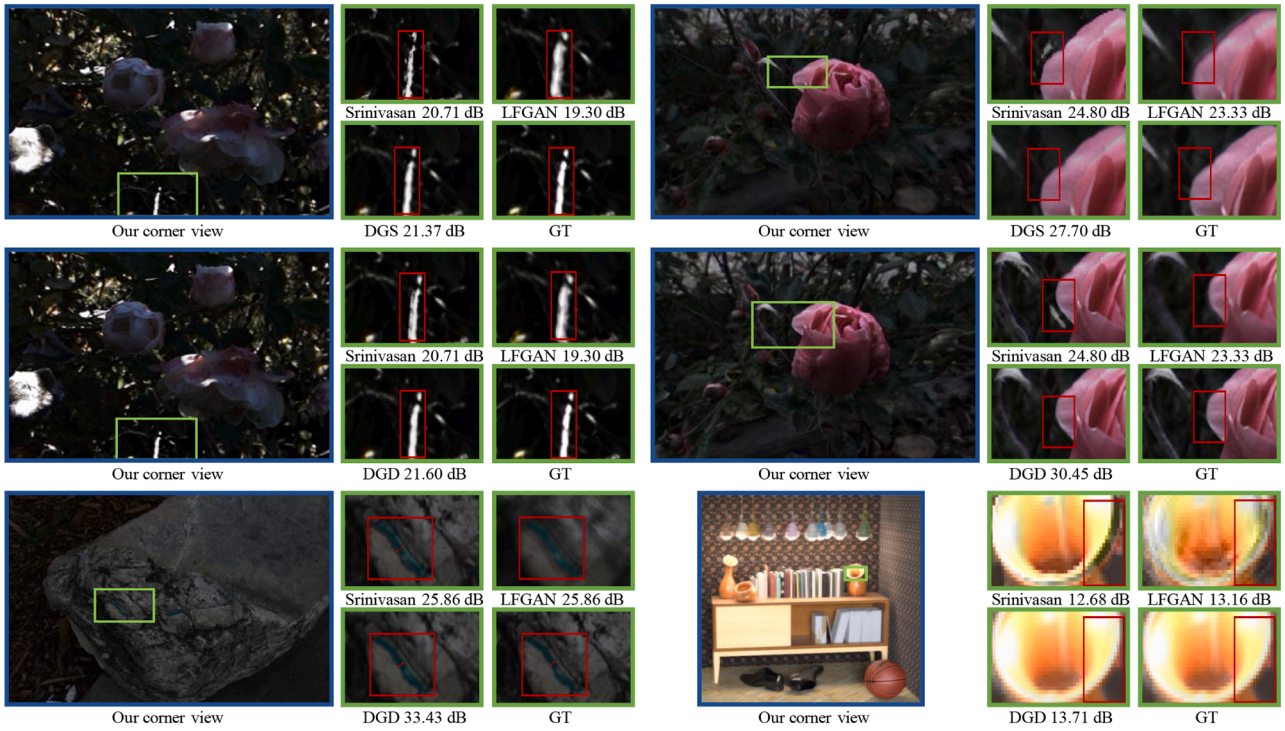


Fig. 4. Qualitative comparison results on Flower, General and HCI dataset. We show the PSNR results generated by each method for a more straightforward comparison. Our DG-GAN generated much more photo realistic results than LFGAN [9] and fewer outliers than [5]. $DGGAN_D$ and $DGGAN_S$ denote DG-GAN using DRN and SRGAN as generator, respectively.

resolutions and a relatively large baseline (distance between cameras) among the five datasets. We randomly select 7 for training and 2 for testing. 5) **Inria** datasets [27], including the Sparse Light Field Dataset (SLFD) with 53 scenes and the Dense Light Field Dataset (DLFD) with 39 scenes, provide more complex scenes than comparable collections. Both datasets offer a resolution of 512×512 with 9×9 angular views, where SLFD features a disparity range of [−20,20] and DLFD ranges between [−4,4].

Compared methods. We compare our proposed method with two categories of state-of-the-arts: 1) geometry-based approaches: Srinivasan et al. [5] and Li et al. [6], ViewCrafter [28], Genwarp [8]; 2) super-resolution-based approaches: LFGAN [9]. We term our proposed method as $DGGAN_S$ and $DGGAN_D$ which exploit SRGAN [13] and DRN [14] as generator, respectively.

Comparisons with single-view view synthesis methods. For a fair comparison, we follow the Single-view View Synthesis methods [29,30], to conduct light field synthesis with $DGGAN_{SRGB}$ (find detailed architecture in section. 3.6) on RGB color space. In testing phase, we use four corner view (i.e.top left, top right, bottom left and bottom right) for calculating each PSNR, SSIM and LPIPS. To further verify our method, we compare our DG-GAN with state-of-the-art methods, including MPI [30], MINE [29], Zhou et al. [31], and LoLep [32], in the context of novel view synthesis from a single-view image. Moreover, we introduce $DGGAN_{SY}$ and $DGGAN_{DY}$, which are two variants of $DGGAN_S$ and $DGGAN_D$ without introducing DGLF.

4.1. Comparisons on local light field synthesis

To evaluate our proposed method, we conduct extensive experiments on five datasets including General, Flower, HCI, Stanford and Inria. In Table 1, our method ($DGGAN_S$ and $DGGAN_D$) achieves the best performance in terms of PSNR and SSIM on all the datasets. From the results, we draw the following observations. 1) Srinivasan et al. [5] yield relatively worse results on most of the datasets, which may be imputed to the negative impact of an inaccurate generation of the disparity map. In other words, geometry-based approaches may be a suboptimal solution in the absence of annotations on disparity maps. Moreover, unlike other methods, Li et al. requires pre-estimated disparity maps (computed by DeepLens [33]) as additional fixed inputs. However, the error of the fixed disparity maps is amplified especially in HCI dataset with complex geometry, resulting in low-quality synthesized light fields with PSNR of only 9.97. We also observe that Li et al. achieves the best SSIM in the General dataset and the best PSNR in the stanford dataset. The main reason is that the two datasets usually do not have complex geometries, so the pre-estimated disparity maps are accurate enough for light field generation. On the flower dataset, with a large amount of data, our DG-GAN also learns a promising geometry and exceeds Li et al. by a large margin via alleviating the side effects of the geometry. 2) LFGAN [9] performs better than Srinivasan et al. [5], which indicates that super-resolution-based approaches without the dependence on the predicted geometries are able to obtain a more accurate estimation on the light field. 3) Compared with the variants $DGGAN_{SY}$ and $DGGAN_{DY}$, our proposed methods ($DGGAN_S$, $DGGAN_D$) achieve promising results. Such results verify the effectiveness of the guidance of disparity maps. That is, our method successfully introduces the geometry and meanwhile mitigate the negative influence caused by the inaccurate estimations of geometry.

4.2. Visualization results of local light fields synthesis

To provide intuitive comparisons between our proposed method and existing state-of-the-arts, we visualize the light field synthesized by them. As shown in Fig. 4, we present 6 groups of comparison results from different datasets. To localize on tiny differences quickly, we zoom in the corner view for better visualization. In the first two columns, we show the generated light field of our method (i.e. $DGGAN_D$ and $DGGAN_S$) using different generator backbones, in the same scenes.

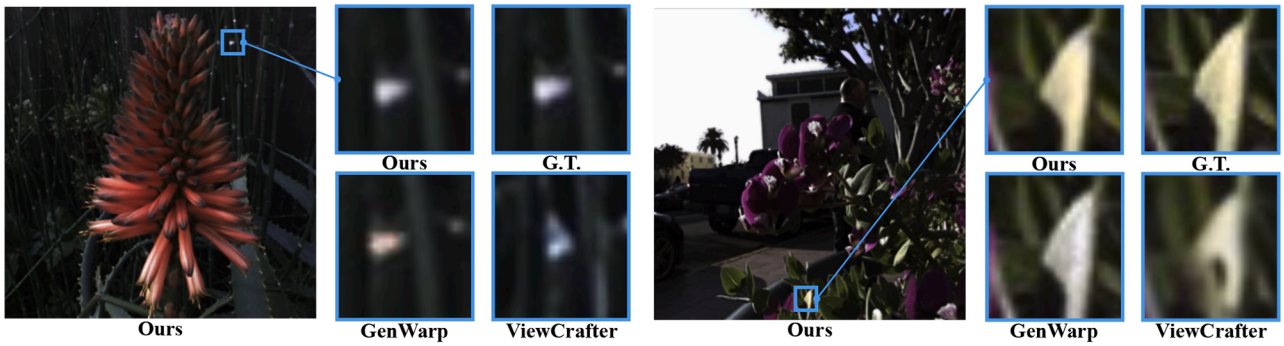


Fig. 5. Comparisons with Genwarp and ViewCrafter, our proposed DG-GAN achieves more accurate color and shape than Genwarp and ViewCrafter.

Table 3
Comparison on depth estimation using light fields generated by different methods on Flower.

Methods	RMSE↓
Srinivasan et al. [5]	7.992
LFGAN [9]	9.646
$DGGAN_D$ (Ours)	**7.833**

Table 4
Comparison with existing methods on Single-view Synthesis using Flower dataset. The best results are in bold font.

Methods	LPIPS↓	SSIM↑	PSNR↑
Li et al. [6]	0.1433	0.857	28.2
MPI [30]	–	0.851	30.1
MINE [29]	0.1559	0.872	30.3
Zhou et al. [31]	0.1582	**0.879**	30.5
LoLep [32]	0.1980	0.868	30.2
$DGGAN_{SRGB}$ (Ours)	**0.1440**	**0.879**	**30.8**

Overall, our method achieves more promising results than LFGAN [9] and Srinivasan et al. [5]. That is, our synthesized light field is closer to the ground truth. In addition, Srinivasan et al. [5] tends to generate more outliers, while LFGAN [9] produces fuzzy results. In the last column, we are able to draw similar observations in these two different scenes. We also compare our method with recent approaches, including ViewCrafter [28] and GenWarp [8], as shown in Fig. 5. These methods leverage diffusion models as image priors and employ off-the-shelf estimators for high-quality geometry extraction. Nevertheless, our method consistently achieves superior color accuracy and shape fidelity across all test datasets.

4.3. Ablation studies and further experiments

In this subsection, we evaluate the influence of the proposed interactive discriminator and quantitatively investigate the geometry information learned by different methods. Moreover, we compare several single-view view synthesis methods to further validate the effectiveness of our DG-GAN.

Impact of interactive discriminator. To study the effect of Inter-discriminator, we compare the perceptual quality of the generated light field obtained from models with different discriminators. In Table 2, compared with DG-GAN with traditional discriminator [34] and without discriminator, our interactive discriminator consistently achieves higher LPIPS on the General datasets. These results show that by exploiting the interaction between angular and spatial domain information, the interactive discriminator enhances the features of the light field data, leading to better performance. Moreover, introducing discriminator into the paradigm obtains a more promising result, which demonstrates that adversarial learning helps to promote the visual reality of the generated light field.

Discussion on learned geometry structure. To investigate the learned geometry information, we quantitatively compare depth maps obtained from the light fields generated by different methods. Following chen et al. [9], the depth maps of the ground truth and the generated light field are obtained by Jeon et al. [35]. As shown in Table 3, our estimated depth maps achieve better results than Srivanisan et al. [5] and LFGAN [9] in terms of RMSE [36]. The result implies that our DG-GAN learns a satisfying geometry structure. We also qualitatively visualize disparity maps directly generated by DG-GAN in Fig. 6.

Comparison with NeRF-based and MPI-based single-view view synthesis methods. To further verify the superiority of DG-GAN, we compare with NeRF-based (MINE [29]) and MPI-based [30] single-view view synthesis (SVVS) methods which seek to synthesize novel views from a single image. Different from us, these methods attempt to synthesize novel views at arbitrary rotation and thus require camera pose and intrinsics as additional input, which are usually estimated in advance by time-consuming structure from motion algorithms. Since the two SVVS methods are designed for RGB color space, we devise a variant, namely $DGGAN_{SRGB}$, to synthesize RGB light fields guided by disparity maps. Following [29] and [30], the training and testing splits are obtained from Tucker and Snavely [30], and the four corner images are the target image for calculating each metric. As shown in Table 4, $DGGAN_{SRGB}$ outperforms MINE [29], MPI [30], li et al. [6], zhou et al. [31] and LoLep [32] regarding LPIPS, SSIM and PSNR. Li et al. uses additional depth information and thus performs well in LPIPS.

4.4. Visualization results on disparity estimation

We qualitatively visualize disparity maps generated by our proposed DG-GAN in Fig. 6. Note that we train the disparity graphs in a self-supervised manner, and therefore do not require disparity labels. From the results, we find that the performance of disparity estimation is correlated with the amount of data trained. Specifically, the flower dataset contains about 3000 light field images, while the General dataset contains about 100, and the general disparity estimation does not perform as well as the flower dataset. This implies that the performance of disparity estimation is expected to be improved by training with more data.

4.5. Visualization results on light field refocusing

We show refocusing capability of our synthesized light field on the Flower and the General datasets. Based on a light field refocusing

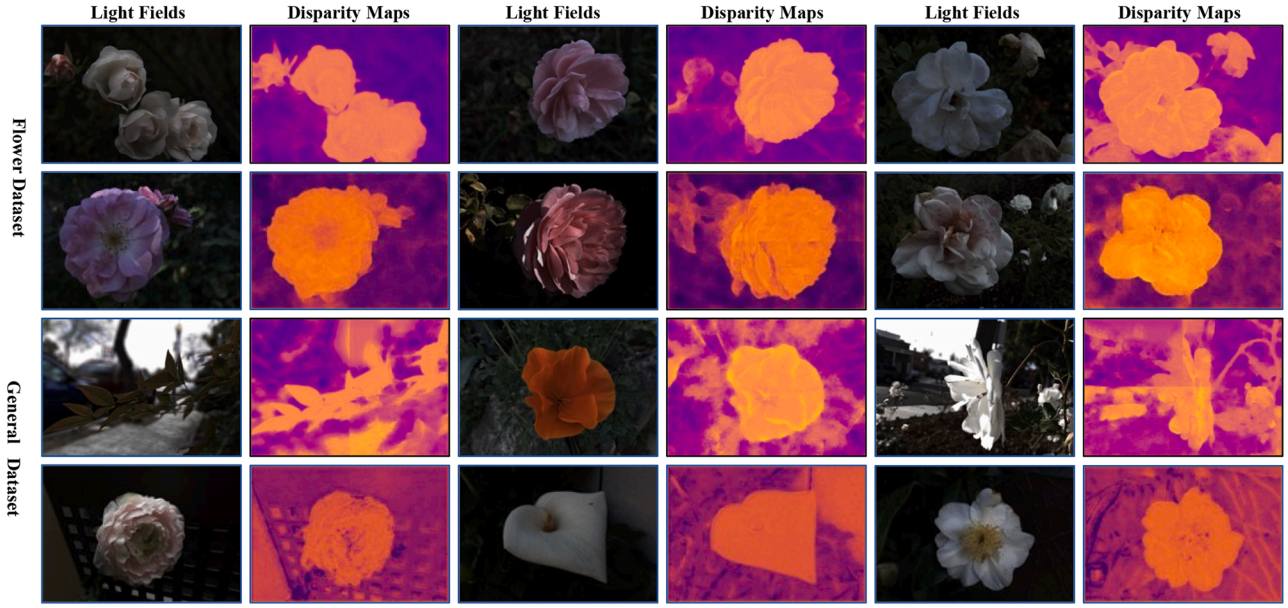


Fig. 6. Demonstration on our estimated disparity map on the Flowers and the General datasets.

Fig. 7. Refocusing results of generated light fields by DG-GAN on the Flowers and the General datasets.

method [37], we refocus generated light field from background to foreground. As shown in Fig. 7, using synthesized light field (first row), we make the area in the background clear and the area in the foreground blurred (second row). Moreover, we are able to make the area in the background blurred and the area in the foreground clear (third row).

4.6. Impact on interactive discriminator

We qualitatively and quantitatively compare the effect of each discriminator I in this section. As shown in Fig. 8, results without using a discriminator and with a traditional discriminator [34] are blurred and contain less texture information. In contrast, our interactive discriminator helps to synthesize light fields with more accurate edges and more detailed texture information. Quantitatively, as demonstrated in Table 5, we find that using a traditional discriminator helps to generate more photo-realistic results (evaluated by LPIPS). Moreover, our proposed interactive discriminator achieves better performance than the traditional discriminator by capturing the interaction between angular information and spatial information of generated light fields.

Table 5
Ablation studies for the discriminators (I) on local light field synthesis (Flower dataset).

w/o I	traditional I	Interactive I	DGLG	LPIPS	PSNR
✓				0.1882	29.03
	✓		✓	0.1531	28.94
		✓		0.1500	28.97
		✓	✓	**0.1482**	**29.11**

Fig. 8. Visualization results generated using different discriminator (I). Details are zoomed.

Table 6
Effect of hyper-parameter λ_0 on the performance of DGGAN$_S$ (General dataset).

λ_0	0.01	0.1	1	10	100
PSNR / SSIM	23.32 / 0.6979	23.37 / 0.7068	**23.50 / 0.7139**	23.35 / 0.7094	23.11 / 0.6797

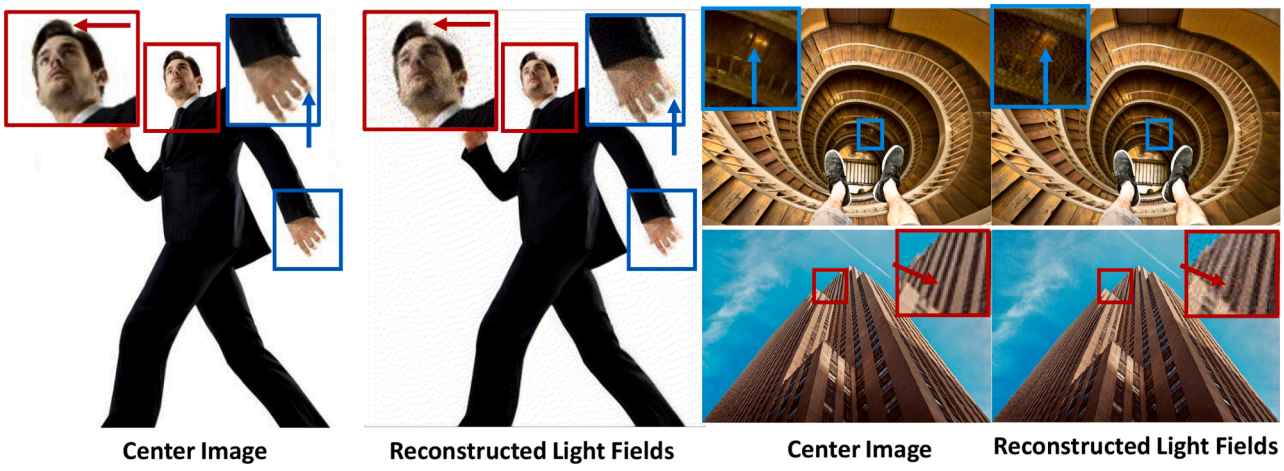


Fig. 9. Failure case of our proposed method on in-the-wild images with strong perspective.

4.7. More discussions on interactive discriminator

In the interactive discriminator, one can choose to sum the spatial feature $\mathbf{p}_1^s$ or angular feature $\mathbf{p}_1^a$ with $\mathbf{p}^{as}$, the output of the bottleneck module, for residual learning. We choose $\mathbf{p}^{as}+\mathbf{p}_1^s$ instead of $\mathbf{p}^{as}+\mathbf{p}_1^a$ because We believe that larger feature size results in stronger representation ability. Specifically, to add $\mathbf{p}_1^a$ or $\mathbf{p}_1^s$ with $\mathbf{p}^{as}$, the size of $\mathbf{p}^{as}$ should be identical to $\mathbf{p}_1^a$ or $\mathbf{p}_1^s$. Additionally, $\mathbf{p}^{as}$ is compressed from $\mathbf{p}_4^a$ and $\mathbf{p}_4^s$ by the bottleneck module. Since the feature map size of $\mathbf{p}_1^s$ is $u\times v$ times larger than $\mathbf{p}_1^a$, we do not need to compress $\mathbf{p}_4^a$ and $\mathbf{p}_4^s$ too much to get $\mathbf{p}^{as}$, resulting in strong representation capability of our proposed Inter-discriminator, see Table 7. Moreover, since we do not need to compress $\mathbf{p}_4^a$ and $\mathbf{p}_4^s$ too much, we can reduce the number of convolutional layers of the bottleneck module, resulting in a more lightweight discriminator than using $\mathbf{p}_1^a+\mathbf{p}^{as}$, see Table 8.

Table 7
Ablation studies on using $\mathbf{p}_1^a+\mathbf{p}^{as}$ and $\mathbf{p}_1^s+\mathbf{p}^{as}$.

Dataset	$\mathbf{p}_1^a+\mathbf{p}^{as}$		$\mathbf{p}_1^s+\mathbf{p}^{as}$	
	PSNR	SSIM	PSNR	SSIM
General	23.44	0.7115	**23.57**	**0.7248**
Inria	**23.95**	**0.6710**	**23.95**	0.6668
Stanford	14.43	0.4050	**14.46**	**0.4099**

Table 8
Comparison of model size for $\mathbf{p}_1^a+\mathbf{p}^{as}$ and $\mathbf{p}_1^s+\mathbf{p}^{as}$.

Model type	#param.
$\mathbf{p}_1^a+\mathbf{p}^{as}$	2.60 M
$\mathbf{p}_1^s+\mathbf{p}^{as}$	**2.53 M**

4.8. Impact of λ_0 in $\mathcal{L}_{DGLF}$

We evaluate our method with different trade-off parameter λ_0 from {0.01, 0.1, 1, 10, 100}. The experiments are conducted by DGGAN$_S$ on the General dataset. From Table 6, we observe that as λ_0 increases from 0.001 to 1, increases from 0.001 to 1, the influence of the disparity map becomes more pronounced, resulting in a progressive enhancement in the performance of light field synthesis. However, when we further increase λ_0 to 10 or 100, the guiding effect of the disparity map overwhelms the primal reconstruction loss $\mathcal{L}_{rec}$, leading to a decline in performance. To achieve an optimal balance between the primal reconstruction and the disparity guidance, we set λ_0 to 1.

4.9. Failure cases of our DG-GAN

Our method performs effectively when handling objects that are present in the training dataset. However, there is room for

Table 9
Architecture of $DGGAN_D$ (the table is from Guo et al. [14]. h and w are height and width of an input image. B is the number of RCABs [38], and F denotes the number of base feature channels. We set $B = 30$ and $F = 8$ for $DGGAN_D$.

Module	Module details	Input shape	Output shape
Head (First Conv)	Conv(3,3)	(1, 8h, 8w)	(1F, 8h, 8w)
Down l	Conv,2-LeakyReLU-Conv	(1F, 8h, 8w)	(2F, 4h, 4w)
Down 2	Conv,2-LeakyReLU-Conv	(2F, 4h, 4w)	(4F, 2h, 2w)
Down 3	Conv,2-LeakyReLU-Conv	(4F, 2h, 2w)	(8F, 1h, 1w)
	B RCABs	(8F, 1h, 1w)	(8F, 1h, 1w)
	2×Upsampler	(8F ,1h, 1w)	(8F, 2h, 2w)
Up 1	Conv(1,1)	(8F ,2h, 2w)	(4F, 2h, 2w)
Concatenation 1	Concatenation of the output of Up 1 and Down 2	(4F, 2h, 2w) ⊕ (4F, 2h, 2w)	(8F, 2h, 2w)
	B RCABs	(8F, 2h, 2w)	(8F, 2h, 2w)
	2×Upsampler	(8F, 2h, 2w)	(8F, 4h, 4w)
Up 2	Conv(1,1)	(8F ,4h, 4w)	(2F, 4h, 4w)
Concatcnation 2	Concatcnation of the output of Up 2 and Down 1	(2F, 4h, 4w) ⊕ (2F, 4h, 4w)	(4F, 4h, 4w)
	B RCABs	(4F ,4h, 4w)	(4F, 4h, 4w)
	2×Upsampler	(4F, 4h, 4w)	(4F, 8h, 8w)
Up 3	Conv(1,1)	(4F, 8h, 8w)	(1F, 8h, 8w)
Concatenation 3	Concatenation of the output of Up3 and Head	(1F, 8h, 8w) ⊕ (1F, 8h, 8w)	(2F, 8h, 8w)
Tail 0	Conv(3,3)	(8F, 1h,1w)	(2, 1h, 1w)
Tail 1	Conv(3,3)	(8F, 2h, 2w)	(2, 2h, 2w)
Tail 2	Conv(3,3)	(4F, 4h, 4w)	(2, 4h, 4w)
Tail 3 (Last Conv)	Conv(3,3)	(2F, 8h, 8w)	(2, 8h, 8w)

improvement in handling in-the-wild scenarios, which are crucial for real-world applications. Additionally, as the majority of light field datasets (including flowers, general scenes) are captured from orthogonal perspectives, methods trained on these datasets struggle to handle images with strong perspective. Consequently, when presented with a single-view image that exhibits a strong perspective, current techniques often fail to generate convincing light field representations. We visualize the performance of our method on in-the-wild images and images with strong perspective in Fig. 9

5. Conclusions and discussion

Conclusion. To reconstruct realistic and geometry-consistent light fields, we propose a disparity-guided generative adversarial network (DG-GAN), which comprises a disparity-guided light field generation (DGLF) paradigm and an interactive discriminator. Our approach mitigates the adverse effects of inaccurate geometry estimation while ensuring the model remains aware of the underlying geometric structure. We provide a theoretical analysis of the generalization performance of the DGLF paradigm and empirically validate its effectiveness. Extensive experiments on five benchmark datasets demonstrate that DG-GAN synthesizes realistic and high-quality geometry-aware light fields, achieving superior performance both quantitatively and qualitatively compared to existing state-of-the-art methods. Through this work, we aim to offer a new perspective on addressing the challenges posed by inaccurate geometry estimation in light field generation.

Future work. We aim to enhance the generalization ability of DG-GAN to reconstruct diverse objects and scenes in real-world scenarios. To achieve this, we plan to integrate large-scale image and video diffusion models, which provide strong real-world priors, with the 3D geometric constraints inherent in light fields. By combining these strengths, we aim to enable in-the-wild novel view synthesis with both realistic appearance and geometrically consistent 3D structures.

Limitations. Our method performs effectively when handling objects that are present in the training dataset. However, there is room for improvement in handling in-the-wild scenarios, which are crucial for real-world applications. Additionally, as the majority of light field datasets (including flowers, general scenes) are captured from orthogonal perspectives, methods trained on these datasets struggle to handle images with strong perspective. Consequently, when presented with a single-view image that exhibits a strong perspective, current techniques often fail to generate convincing light field representations.

CRediT authorship contribution statement

Yifan Yang: Writing – review & editing, Writing – original draft, Visualization, Validation, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization; **Zhen Qiu:** Writing – review & editing, Writing – original draft, Conceptualization; **Shuhai Zhang:** Writing – review & editing, Writing – original draft, Formal analysis; **Mingkui Tan:** Writing – review & editing, Validation, Supervision, Project administration, Investigation, Funding acquisition, Formal analysis, Data curation.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work the authors used ChatGPT in order to correct grammar. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Data availability

Data will be made available on request.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This work was partially supported by the Joint Funds of the National Natural Science Foundation of China (Grant No. U24A20327),

National Natural Science Foundation of China (NSFC) 62072190, Program for Guangdong Introducing Innovative and Entrepreneurial Teams 2017ZT07X183, and TCL Science and Technology Innovation Fund.

In the **supplementary**, we provide the detailed proof of Theorem 1, more details and more experimental results of the proposed DG-GAN. We organize the supplementary material as follows: (1) In PPT file of the supplementary material, we provide synthesized light fields in GIF format. (2) In Appendix A, we derive the generalization error bound of our proposed DGLF.

Appendix A. Proof of Theorem 1

To analyze the generalization performance for the proposed disparity-guided light field (DGLF) generator, we first define the Rademacher complexity of our DGLF scheme.

Definition 1 (Rademacher complexity of DGLF)**.** Given an underlying distribution $\mathcal{P}$ and its empirical distribution $\mathcal{Z}=\{\mathbf{z}_1,\mathbf{z}_2,\dots,\mathbf{z}_N\}$, where $\mathbf{z}_i=(\mathbf{S}(\mathbf{x},\mathbf{0})_i,\mathbf{S}_i)$. Then the Rademacher complexity of DGLF is defined below:

$$\mathrm{R}_N(\mathcal{H}^{DGLF})=\mathbb{E}_{\mathcal{S}\sim\mathcal{P}^N}\left[\hat{\mathrm{R}}_{\mathcal{S}}(\mathcal{H}^{DGLF})\right],\ \text{s.t. } \mathcal{H}^{DGLF}\in\mathcal{H}_f\times\mathcal{H}_{g_S}\times\mathcal{H}_{g_D}. \tag{A.1}$$

where $\mathcal{H}_f$, $\mathcal{H}_{g_S}$, $\mathcal{H}_{g_D}$ are function spaces that are determined by f,g_S,g_D, respectively, $\hat{\mathrm{R}}_{\mathcal{S}}(\mathcal{H}^{DGLF})$ is the empirical Rademacher complexity, which is defined as:

$$\hat{\mathrm{R}}_{\mathcal{S}}(\mathcal{H}^{DGLF})=\mathbb{E}_{\sigma}\left[\sup\frac{1}{N}\sum_{i=1}^{N}\sigma_i\left(\mathcal{L}_{rec}(\mathbf{x},\mathbf{u})+\lambda_0\mathcal{L}_{guide}(\mathbf{x},\mathbf{u})\Big|_{\mathbf{z}_i}\right)\right]. \tag{A.2}$$

where $\sigma=[\sigma_1,\dots,\sigma_N]$ are independent uniform random variables valued in $\{-1,+1\}$.

Recall that $E(f,g_S,g_D)=\mathbb{E}[\mathcal{L}_{rec}(\mathbf{x},\mathbf{u})+\lambda_0\mathcal{L}_{guide}(\mathbf{x},\mathbf{u})]$ and $\hat{E}(f,g_S,g_D)$ is its empirical loss, based on the definition (1), we restate the following theorem for the generalization performance of DGLF.

Theorem 1 (Generalization performance of DGLF) *Let $\mathcal{L}_{rec}(\mathbf{x},\mathbf{u})+\lambda_0\mathcal{L}_{guide}(\mathbf{x},\mathbf{u})$ be a mapping from $S^{center}\times S^{gt}$ to $[0,B]$ with upper bound B and function space $\mathcal{H}^{DGLF}$. For any $\delta>0$, all $(f,g_S,g_D)\in\mathcal{H}^{DGLF}$, with probability at least $1-\delta$, the generalization error $E(f,g_S,g_D)$, i.e., the expected loss, satisfies*

$$E(f,g_S,g_D)\le\hat{E}(f,g_S,g_D)+2\mathrm{R}_N(\mathcal{H}^{DGLF})+B\sqrt{\frac{1}{2N}\log\left(\frac{1}{\delta}\right)}. \tag{A.3}$$

where $\mathrm{R}_N(\mathcal{H}^{DGLF})$ is the Rademacher complexity of our proposed disparity-guided light field generator. Let $\mathcal{B}(f,g_S,g_D)$ be the generalization bound w.r.t. $\mathrm{R}_N(\mathcal{H}^{DGLF})$, i.e., $\mathcal{B}(f,g_S,g_D)=2\mathrm{R}_N(\mathcal{H}^{DGLF})+B\sqrt{\frac{1}{2N}\log\left(\frac{1}{\delta}\right)}$, then we have

$$\mathcal{B}(f,g_S,g_D)\le\mathcal{B}(f,g_S), \tag{A.4}$$

where $(f,g_S)\in\mathcal{H}^{LF}$, $\mathcal{B}(f,g_S)$ is the generalization bound without disparity guide (LF) paradigm w.r.t. the Rademacher complexity $\mathrm{R}_N(\mathcal{H}^{LF})$.

Proof. For any sample $\mathcal{Z}=\{\mathbf{z}_1,\mathbf{z}_2,\dots,\mathbf{z}_N\}$ and any $(f,g_S,g_D)\in\mathcal{H}^{DGLF}$, let $\hat{E}_{\mathcal{Z}}(f,g_S,g_D)$ be the empirical loss of $\mathcal{L}_{DGLF}$ over $\mathcal{Z}$ and define a function Φ for sample $\mathcal{Z}$ as below:

$$\Phi(\mathcal{Z})=\sup_{(f,g_S,g_D)\in\mathcal{H}^{DGLF}}E(f,g_S,g_D)-\hat{E}_{\mathcal{Z}}(f,g_S,g_D). \tag{A.5}$$

Based on McDiarmid's inequality [39] and Theorem 3.1 in Mohri et al. [23], for any $\delta>0$, with probability at least $1-\delta$, we have

$$\Phi(\mathcal{Z})\le\mathbb{E}_{\mathcal{Z}}[\Phi(\mathcal{Z})]+B\sqrt{\frac{1}{2N}\log\left(\frac{1}{\delta}\right)}\le 2\mathrm{R}_N(\mathcal{H}^{DGLF})+B\sqrt{\frac{1}{2N}\log\left(\frac{1}{\delta}\right)}. \tag{A.6}$$

Combining Eq. (A.5) and Inequalities (A.6), with probability at least $1-\delta$, we have

$$E(f,g_S,g_D)\le\hat{E}(f,g_S,g_D)+2\mathrm{R}_N(\mathcal{H}^{DGLF})+B\sqrt{\frac{1}{2N}\log\left(\frac{1}{\delta}\right)}. \tag{A.7}$$

Based on Rademacher complexity, since the capacity of the function space $\mathcal{H}^{DGLF}\in\mathcal{H}_f\times\mathcal{H}_{g_S}\times\mathcal{H}_{g_D}$ is smaller than the capacity of the function space $\mathcal{H}^{LF}\in\mathcal{H}_f\times\mathcal{H}_{g_S}$, we have

$$\mathrm{R}_N(\mathcal{H}^{DGLF})\le\mathrm{R}_N(\mathcal{H}^{LF}). \tag{A.8}$$

With the same number of training samples, we have

$$\mathcal{B}(f,g_S,g_D)\le\mathcal{B}(f,g_S). \tag{A.9}$$

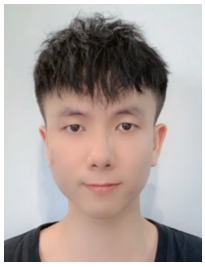
□

References

[1] A. Gershun, The light field, J. Math. Phys. 18 (1–4) (1939) 51–151.
[2] D. Kim, S.-W. Nam, S. Choi, J.-M. Seo, G. Wetzstein, Y. Jeong, Holographic parallax improves 3D perceptual realism, ACM Trans. Graph. 43 (2024) 13.
[3] C. Zhang, G. Hou, Z. Zhang, Z. Sun, T. Tan, Efficient auto-refocusing for light field camera, Pattern Recognit. 81 (2018) 176–189.
[4] Y. Liu, M. Aleksandrov, Z. Hu, Y. Meng, L. Zhang, S. Zlatanova, H. Ai, P. Tao, Accurate light field depth estimation under occlusion, Pattern Recognit. 138 (2023) 109415.
[5] P.P. Srinivasan, T. Wang, A. Sreelal, R. Ramamoorthi, R. Ng, Learning to synthesize a 4D RGBD light field from a single image, in: Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 2243–2251.
[6] Q. Li, N.K. Kalantari, Synthesizing light field from a single image with variable MPI and two network fusion, ACM Trans. Graph. 39 (6) (2020) 229–1.
[7] J. Bak, I. Kyu Park, Light field synthesis from a monocular image using variable LDI, in: 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2023, pp. 3399–3407.
[8] J. Seo, K. Fukuda, T. Shibuya, T. Narihira, N. Murata, S. Hu, C.-H. Lai, S. Kim, Y. Mitsufuji, Genwarp: single image to novel views with semantic-preserving generative warping, in: The Thirty-eighth Annual Conference on Neural Information Processing Systems, 2024.
[9] B. Chen, L. Ruan, M.-L. Lam, LFGAN: 4D Light field synthesis from a single RGB image, ACM Trans. Multimed. Comput. Commun. Appl. 16 (1) (2020) 20.
[10] P. Chandramouli, K.V. Gandikota, A. Goerlitz, A. Kolb, M. Moeller, A generative model for generic light field reconstruction, IEEE Trans. Pattern Anal. Mach. Intell. 44 (4) (2022) 1712–1724.
[11] J. Jin, J. Hou, H. Yuan, S. Kwong, Learning light field angular super-resolution via a geometry-aware network, in: Proceedings of the AAAI Conference on Artificial Intelligence, 34, 2020, pp. 11141–11148.
[12] Y. Wang, L. Wang, J. Yang, W. An, J. Yu, Y. Guo, Spatial-angular interaction for light field image super-resolution, in: European Conference on Computer Vision, Springer, 2020, pp. 290–308.
[13] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, et al., Photo-realistic single image super-resolution using a generative adversarial network, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2017, pp. 4681–4690.
[14] Y. Guo, J. Chen, J. Wang, Q. Chen, J. Cao, Z. Deng, Y. Xu, M. Tan, Closed-loop matters: dual regression networks for single image super-resolution, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 5407–5416.
[15] N.K. Kalantari, T.-C. Wang, R. Ramamoorthi, Learning-based view synthesis for light field cameras, ACM Trans. Graph. 35 (6) (2016) 193:1–193:10.
[16] Y. Li, X. Wang, H. Zhu, G. Zhou, Q. Wang, Dense light field reconstruction based on epipolar focus spectrum, Pattern Recognit. 140 (2023) 109551.
[17] Y. Liao, L. Song, G. Zhang, F. Fang, Multi-level disparity-guided transformers for light field spatial super-resolution, Pattern Recognit. 168 (2025) 111803.
[18] J. Peng, Z. Xiong, D. Liu, X. Chen, Unsupervised depth estimation from light field using a convolutional neural network, in: 2018 International Conference on 3D Vision (3DV), 2018, pp. 295–303.
[19] Y. Yoon, H.-G. Jeon, D. Yoo, J.-Y. Lee, I. So Kweon, Learning a deep convolutional network for light-field image super-resolution, in: Proceedings of the IEEE International Conference on Computer Vision Workshops, 2015, pp. 24–32.
[20] M.S.K. Gul, B.K. Gunturk, Spatial and angular resolution enhancement of light fields using convolutional neural networks, IEEE Trans. Image Process. 27 (5) (2018) 2146–2159.
[21] Y. Mo, Y. Wang, C. Xiao, J. Yang, W. An, Dense dual-Attention network for light field image super-resolution, IEEE Trans. Circuits Syst. Video Technol. 32 (7) (2022) 4431–4443.
[22] G. Wu, M. Zhao, L. Wang, Q. Dai, T. Chai, Y. Liu, Light field reconstruction using deep convolutional network on EPI, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2017, pp. 6319–6327.
[23] M. Mohri, A. Rostamizadeh, A. Talwalkar, Foundations of Machine Learning, MIT Press, 2012.
[24] Y. Wang, L. Wang, J. Yang, W. An, J. Yu, Y. Guo, Spatial-angular interaction for light field image super-resolution, in: European Conference on Computer Vision (ECCV), 2020, pp. 290–308.

[25] K. Honauer, O. Johannsen, D. Kondermann, B. Goldluecke, A dataset and evaluation methodology for depth estimation on 4D light fields, in: Asian Conference on Computer Vision, Springer, 2016, pp. 19–34.
[26] M. Rerabek, T. Ebrahimi, New light field image dataset, in: 8th International Conference on Quality of Multimedia Experience (QoMEX), CONF, 2016.
[27] J. Shi, X. Jiang, C. Guillemot, A framework for learning depth from a flexible subset of dense and sparse light field views, IEEE Trans. Image Process. 28 (12) (2019) 5867–5880.
[28] W. Yu, J. Xing, L. Yuan, W. Hu, X. Li, Z. Huang, X. Gao, T.-T. Wong, Y. Shan, Y. Tian, Viewcrafter: taming video diffusion models for high-fidelity novel view synthesis, arXiv preprint arXiv:2409.02048 0 (2024) 0.
[29] J. Li, Z. Feng, Q. She, H. Ding, C. Wang, G.H. Lee, Mine: towards continuous depth mpi with nerf for novel view synthesis, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 12578–12588.
[30] R. Tucker, N. Snavely, Single-view view synthesis with multiplane images, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 551–560.
[31] Y. Zhou, H. Wu, W. Liu, Z. Xiong, J. Qin, S. He, Single-view view synthesis with self-rectified pseudo-stereo, Int. J. Comput. Vis. 131 (8) (2023) 2032–2043.
[32] C. Wang, Y.-P. Wang, D. Manocha, LoLep: single-view view synthesis with locally-learned planes and self-attention occlusion inference, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2023, pp. 10841–10851.
[33] W. Lijun, S. Xiaohui, Z. Jianming, W. Oliver, L. Zhe, H. Chih-Yao, K. Sarah, L. Huchuan, DeepLens: shallow depth of field from a single image, ACM Trans. Graph. (Proc. SIGGRAPH Asia) 37 (6) (2018) 6:1–6:11.
[34] A. Radford, L. Metz, S. Chintala, Unsupervised representation learning with deep convolutional generative adversarial networks, International Conference on Learning Representations (2016) 0.
[35] H.-G. Jeon, J. Park, G. Choe, J. Park, Y. Bok, Y.-W. Tai, I. So Kweon, Accurate depth map estimation from a lenslet light field camera, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2015, pp. 1547–1555.
[36] C. Godard, O. Mac Aodha, G.J. Brostow, Unsupervised monocular depth estimation with left-right consistency, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2017.
[37] Y. Wang, J. Yang, Y. Guo, C. Xiao, W. An, Selective light field refocusing for camera arrays using bokeh rendering and superresolution, IEEE Signal Process. Lett. 26 (2019) 204–208.
[38] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, Y. Fu, Image super-resolution using very deep residual channel attention networks, in: Proceedings of the European Conference on Computer Vision, 2018, pp. 286–301.
[39] W. Hoeffding, Probability inequalities for sums of bounded random variables, The Collected Works of Wassily Hoeffding (1994) 409–426.

Yifan Yang received the Ph.D. degree in the school of computer science and engineering from South China University of Technology, Guangzhou, China. His research interests include 3D reconstruction, digital human, and medical imaging.



Zhen Qiu received the bachelor's degree in information engineering and the master's degree in software engineering from South China University of Technology, Guangzhou, China, in 2019 and 2022, respectively. His research interests include deep learning and computer vision.

Shuhai Zhang received the bachelor's degree in software engineering from the School of Software Engineering, South China University of Technology Guangzhou, China in 2020, where he is currently pursuing the Ph.D degree. His research interests include generative models and adversarial learning.

Mingkui Tan received the bachelor's degree in environmental science and engineering and the master's degree in control science and engineering from Hunan University, Changsha, China, in 2006 and 2009, respectively, and the Ph.D. degree in computer science from Nanyang Technological University, Singapore, in 2014. From 2014 to 2016, he has worked as a Senior Research Associate on computer vision with the School of Computer Science, The University of Adelaide, Australia. He is currently a Professor with the School of Software Engineering, South China University of Technology. His research interests include machine learning, sparse analysis, deep learning, and large-scale optimization.