LIGHT-FIELD RECONSTRUCTION AND DEPTH ESTIMATION FROM FOCAL STACK IMAGES USING CONVOLUTIONAL NEURAL NETWORKS

Zhengyu Huang[†], *Jeffrey A. Fessler*[†], *Theodore B. Norris*[†], *Il Yong Chun*[⋆]

[†] Department of EECS, University of Michigan, Ann Arbor, MI, USA
[⋆] Department of EE, University of Hawai'i at Mānoa, HI, USA

ABSTRACT

Light-field (LF) reconstruction from focal stack images has diverse applications including face recognition, autonomous driving, and 3D reconstruction in virtual reality. It is a large-scale ill-conditioned inverse problem and typically requires regularized iterative algorithms to solve, which can be slow. This paper proposes a *non-iterative* LF reconstruction and depth estimation method based on three sequential convolutional neural networks (CNNs). The first CNN estimates an all-in-focus image from focal stack images. The second CNN estimates 4D ray depth from the estimated all-in-focus image via the first CNN, and focal stack images. The third CNN refines a Lambertian LF that is rendered using the all-in-focus image and ray depth estimated by the first and second CNNs, respectively. Numerical experiments show that the proposed CNN-based method achieves significantly more accurate and/or faster LF reconstruction, compared to a state-of-the-art sequential CNN using a single image, conventional model-based image reconstruction from a focal stack, and direct regression CNN from a focal stack.

***Index Terms*—** Light-field reconstruction, Depth estimation, Focal stack, Inverse problem, Neural network

1. INTRODUCTION

In conventional photography, a pixel value is formed by integrating light rays coming from different directions and hence the directional information is lost. On the other hand, the directional information is preserved in light-field (LF) photography, because a 4D LF records the directional distribution of the light ray passing through each location of a 2D plane. 4D LFs lead to better scene understanding, such as more accurate depth estimation, better object detection and recognition [1, 2, 3, 4]; 4D LF photography has numerous applications including face recognition, autonomous driving, and 3D reconstruction in virtual reality.

A 4D LF can be parameterized using two parallel reference planes placed at arbitrary positions. In this two-plane parameterization, every light ray can be identified by its interception at the first plane coordinates $\boldsymbol{\nu} = (u, v)$ and the second plane coordinates $\boldsymbol{x} = (s, t)$, with its radiance $l(\boldsymbol{x}, \boldsymbol{\nu})$. There are multiple approaches for directly acquiring 4D LFs. Examples include a 2D camera array, a camera mounted on a 2D gantry with sequential exposures at different positions, and plenoptic cameras, e.g., Lytro and RayTrix. However, these methods have limitations: a 2D camera array is expensive; for single camera on 2D gantry, the number of exposures increase, as the angular resolution of LFs increases; plenoptic cameras have a trade-off between the spatial and angular resolutions in 4D LFs.

To overcome the aforementioned limitations, researchers have proposed model-based image reconstruction (MBIR) methods that reconstruct a 4D LF from a set of limited focal stack images [5, 6, 7, 8]. Reconstructing a 4D LF from a 3D focal stack, i.e., a set of 2D images captured by sensor at multiple different locations along the optical axis, can be viewed as solving a large-scale ill-posed inverse problem; sophisticated regularizer designs can improve reconstruction quality. Assuming a Lambertian LF, Levin and Durand used a 3D Gaussian prior to reconstruct a LF from a focal stack [5]. Blocker et al. proposed a MBIR method that uses a low rank plus sparse tensor regularizer [8]. Since these MBIR methods are iterative, they become slower as one attempts to recover higher-resolution LFs.

Convolutional neural network (CNN) methods are rapidly emerging as a powerful tool for various image processing and computer vision tasks due to their ability to model complicated functions and short inference time [9]. CNN methods also have been applied to inverse imaging problems, including computed tomography [10], magnetic resonance imaging [11], positron emission tomography [12], and LF view synthesis [13, 14, 15]. Srinivasan et al. proposed a sequential CNN approach that reconstructs a LF from a single all-in-focus image [14]. Their pipeline consists of a CNN that estimates ray depth of the scene from the input all-in-focus image, a rendering module that renders a Lambertian LF using the estimated ray depth, and a second CNN that corrects the artifacts in the rendered Lambertian LF. As their method uses ray depth to render a LF, LF reconstruction quality largely depends on the quality of the estimated ray depth. However, depth estimation from single image is challenging as it lacks

This work was supported in part by the Keck Foundation.

978-1-5090-6631-5/20/$31.00 ©2020 IEEE

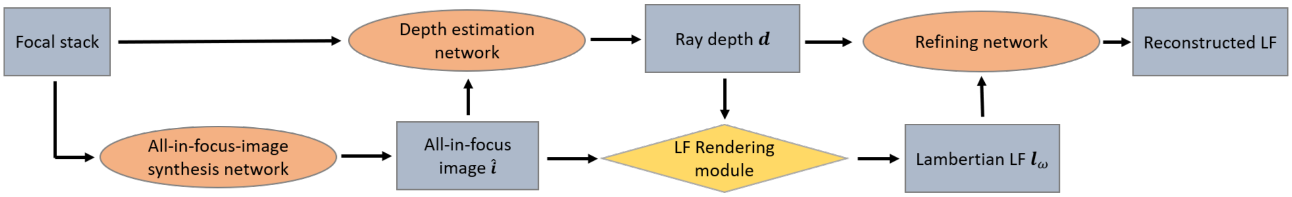
Focal stack
Depth estimation network
Ray depth $\boldsymbol{d}$
Refining network
Reconstructed LF
All-in-focus-image synthesis network
All-in-focus image $\hat{\boldsymbol{i}}$
LF Rendering module
Lambertian LF $\boldsymbol{l}_\omega$

Fig. 1. Proposed CNN-based method for LF reconstruction and depth estimation using focal stack.

reliable depth cues. As a result, a better depth estimation would greatly benefit LF reconstruction. Motivated by this, [15] used coded projection images (at most two) to obtain more accurate ray depth and LF, by modifying the pipeline in [14].

This paper proposes a sequential CNN framework that reconstructs a LF and estimates a depth map from focal stack images, instead of from a single all-in-focus image or coded projection images. A focal stack consists of a set of images focused at different depths of the scene. In each image, only the part of the scene that is in-focus is sharp and out-of-focus regions are blurred by depth-dependent amounts. This diversity provides useful depth cues and can be exploited to improve depth estimation and LF reconstruction. Motivated by [15], our proposed method uses three sequential CNNs. The first CNN estimates an all-in-focus image from focal stack images; the second CNN estimates 4D ray depth from focal stack images and the estimated all-in-focus image; a rendering module renders a Lambertian LF with the estimated all-in-focus image and ray depth, and the third CNN subsequently refines the rendered LF and provides the final reconstructed LF. Numerical experiments show that the proposed method significantly improves LF reconstruction accuracy, compared with a state-of-the-art sequential CNN approach using a single all-in-focus image [14], conventional MBIR using 4D edge-preserving (EP) regularizer (from a focal stack) [8], and direct regression CNN from a focal stack. In addition, the proposed method considerably reduces LF reconstruction time compared with MBIR using EP regularizer.

2. METHOD

The proposed approach uses four steps to reconstruct LFs, as illustrated in Fig. 1. In the first step, an "all-in-focus image synthesis" neural network (NN) synthesizes an all-in-focus image from a focal stack (Section 2.1). In the second step, a depth estimation NN estimates 4D ray depth $\boldsymbol{d}$ (depth maps for every view point) from the estimated all-in-focus image $\hat{\boldsymbol{i}}$ and focal stack images $\boldsymbol{f}$ (Section 2.2). The third step renders a Lambertian LF $\boldsymbol{l}_w$ by backward warping the all-in-focus image $\hat{\boldsymbol{i}}$, using the estimated 4D ray depth $\boldsymbol{d}$ (Section 2.3). Because the rendered LF is Lambertian and may contain artifacts around occlusions, we use a refining NN to further refine $\boldsymbol{l}_w$ and obtain a final LF $\boldsymbol{l}^\star$ (Section 2.4). The following subsections describe details of each step.

2.1. All-in-focus image synthesis NN

We first estimate an all-in-focus image, given the focal stack images; this process is called focal stacking. There are several focal stacking approaches, e.g., edge detection, Fourier analysis, and CNN. Among these, we choose CNN-based method – specifically, U-Net [16] with modified input and output channel numbers – due to its good image mapping capability. We forward pass reshaped focal stack images (from the size $C \times N_F \times H \times W$ to $(C \cdot N_F) \times H \times W$, where C is the number of color channel, N_F is the number of focal planes in the focal stack, H and W are the image height and width, respectively) through the modified U-Net. To squeeze the output all-in-focus image to be within the interval $[0, 1]$, we put a differentiable nonlinear function $g(\cdot) = (\tanh(\cdot) + 1)/2$ at the end of the U-Net. We train the modified U-Net, $\mathcal{A}_{\theta_a}(\boldsymbol{f})$, having parameter set θ_a, by minimizing the ℓ_1 loss:

$$\min_{\theta_a} \sum_n \|\mathcal{A}_{\theta_a}(\boldsymbol{f}_n) - \boldsymbol{i}_n\|_1,$$

where $\{\boldsymbol{f}_n : \forall n\}$ are training focal stack images, and $\{\boldsymbol{i}_n : \forall n\}$ are the ground truth all-in-focus images. We use the center sub-aperture images of the ground truth LFs for $\{\boldsymbol{i}_n\}$, because sub-aperture images of LFs have small enough aperture such that all regions of the image are well in focus.

2.2. Depth estimation NN

The LF rendering (Section 2.3 uses both an all-in-focus image and a 4D ray depth $d(\boldsymbol{x}, \boldsymbol{\nu})$, i.e., a collection of 2D disparity maps, one for each angular coordinate $\boldsymbol{\nu}$. We modify the CNN architecture in [14] to estimate 4D ray depth using focal stack images and the all-in-focus image from $\mathcal{A}_{\theta_a}$. We reshape the input focal stack as in Section 2.1). We use dilated convolution layers [17] to have exponentially growing receptive field without losing resolution. At the end of the NN, a $\tanh$ scaling layer squeezes the estimated disparity within the range $[-1, 1]$. We jointly train the depth estimation NN and the refininig NN (Section 2.4); see training loss in Section 2.5.

2.3. Light field rendering

Given the estimated 4D ray depth $\boldsymbol{d}$ and the estimated all-in-focus image $\hat{\boldsymbol{i}}$ via trained $\mathcal{A}_{\theta_a^\star}(\cdot)$, we render a Lambertian LF

$\boldsymbol{l}_w$ by backward warping $\hat{\boldsymbol{i}}$ as follows [14]:

$$
\begin{aligned}
l_w(\boldsymbol{x}, \boldsymbol{\nu}) = l_w(\boldsymbol{x} + \boldsymbol{\nu} d(\boldsymbol{x}, \boldsymbol{\nu}), \mathbf{0}) = \hat{i}(\boldsymbol{x} + \boldsymbol{\nu} d(\boldsymbol{x}, \boldsymbol{\nu})) \\
=: \mathcal{W}(\hat{\boldsymbol{i}}, \boldsymbol{d}).
\end{aligned} \tag{1}
$$

We use bilinear interpolation to calculate the values of $\hat{i}(\boldsymbol{x} + \boldsymbol{\nu} d(\boldsymbol{x}, \boldsymbol{\nu}))$ in the warping. As the rendering at a viewpoint $\boldsymbol{\nu}$ given by (1) is essentially a sampling of the pixel values at the center view, the rendered LF $\boldsymbol{l}_w$ will be approximately Lambertian and can have artifacts around the occlusion regions.

2.4. Refining NN

Because the rendered LF $\boldsymbol{l}_w$ from Section 2.3 does not model the non-Lambertian effect and occlusion effect, we use an additional refining NN (see its architecture in [14]) to remove these artifacts and get a final reconstructed LF $\hat{\boldsymbol{l}}$. We use a residual connection [18] for the NN to learn the difference between the Lambertian LF $\boldsymbol{l}_w$ and true LF $\boldsymbol{l}$. We input both estimated 4D ray depth $\boldsymbol{d}$ and Lambertian LF $\boldsymbol{l}_w$ to the NN; in particular, $\boldsymbol{d}$ is useful for predicting occluded region and to refine $\boldsymbol{l}_w$.

2.5. Training of depth estimation NN and refining NN

We jointly train the depth estimation NN and the refining NN, similar to [14]. By using differentiable bilinear interpolation for the LF rendering, the loss gradient can be backpropagated from the refining NN, through the LF rendering module, and to the depth estimation NN. Specifically, we jointly train a depth estimation NN, $\mathcal{D}_{\theta_d}(\boldsymbol{f}, \hat{\boldsymbol{i}})$, and a refining NN, $\mathcal{R}_{\theta_r}(\boldsymbol{d}, \boldsymbol{l}_w)$ having parameters θ_d and θ_r, respectively, by minimizing the following loss function:

$$
\begin{aligned}
\min_{\theta_d, \theta_r} \sum_n \left\| \mathcal{W}\big(\hat{\boldsymbol{i}}_n, \mathcal{D}_{\theta_d}(\boldsymbol{f}_n, \hat{\boldsymbol{i}}_n)\big) - \boldsymbol{l}_n \right\|_1 + \\
\left\| \mathcal{R}_{\theta_r}\Big(\mathcal{D}_{\theta_d}(\boldsymbol{f}_\theta, \hat{\boldsymbol{i}}_n), \mathcal{W}\big(\hat{\boldsymbol{i}}_n, \mathcal{D}_{\theta_d}(\boldsymbol{f}_n, \hat{\boldsymbol{i}}_n)\big)\Big) - \boldsymbol{l}_n \right\|_1 + \\
\lambda_{\mathrm{c}} \psi_{\mathrm{c}}\Big(\mathcal{D}_{\theta_d}(\boldsymbol{f}_n, \hat{\boldsymbol{i}}_n)\Big) + \lambda_{\mathrm{tv}} \psi_{\mathrm{tv}}\Big(\mathcal{D}_{\theta_d}(\boldsymbol{f}_n, \hat{\boldsymbol{i}}_n)\Big),
\end{aligned} \tag{2}
$$

where the training data consists of focal stack images $\{\boldsymbol{f}_n\}$, estimated all-in-focus images $\{\hat{\boldsymbol{i}}_n = \mathcal{A}_{\theta_a^\star}(\boldsymbol{f}_n) : \forall n\}$ and ground truth LFs $\{\boldsymbol{l}_n : \forall n\}$. In (2), ψ_{c} and ψ_{tv} are 4D ray depth consistency and total variation regularizer, respectively, designed to make the estimated ray depth $\boldsymbol{d}$ reasonable [14]. The regularizers are defined by [14]

$$
\psi_{\mathrm{c}}(\boldsymbol{d}) := \sum_{\boldsymbol{x}, \boldsymbol{\nu}} |d(\boldsymbol{x}, \boldsymbol{\nu}) - d(\boldsymbol{x} + d(\boldsymbol{x}, \boldsymbol{\nu}), \boldsymbol{\nu} - \mathbf{1})| \tag{3}
$$

$$
\psi_{\mathrm{tv}}(\boldsymbol{d}) := \|\nabla_{\boldsymbol{x}} \boldsymbol{d}\|_1. \tag{4}
$$

As the ray depth consists of depth maps at different viewpoints, these depth maps should be consistent with each other. Specifically, they are related by the equality

$$
d(\boldsymbol{x}, \boldsymbol{\nu}) = d(\boldsymbol{x} + \boldsymbol{\Delta} D(\boldsymbol{x}, \boldsymbol{\nu}), \boldsymbol{\nu} - \boldsymbol{\Delta}) \tag{5}
$$

that is similar to the relation in (1). Note that the relation in (1) corresponds to a special case of $\boldsymbol{\Delta} = \boldsymbol{\nu}$; choosing $\boldsymbol{\Delta} = \mathbf{1}$ leads to the ray depth consistency regularizer in (3), which encourages depths maps at neighboring views to be consistent. On the other hand, the total variation regularizer in (4) ensures the estimated depth maps are spatially smooth.

3. EXPERIMENTAL SETUP

We compared the proposed method with following three methods: *1)* a state-of-the-art sequential CNN method that estimates 4D ray depth from a single image and then reconstructs a LF [14]; *2)* a conventional 4D EP MBIR method that reconstructs a LF from focal stack (see, e.g., [8, 19]); *3)* a direct regression CNN from focal stack – we chose a U-Net architecture [16]. For *3)*, a sufficient number of network parameters is chosen such that further increasing the parameter doesn't give better performance.

3.1. Dataset and imaging simulation

For all experiments in the paper, we used the LF dataset in [14] that consists of 3343 RGB LFs of flowers and plants taken with Lytro Illum camera. To avoid an inverse crime, we simulated 185×269 focal stack images with number of focal planes $N_F = 7$, from high spatial resolution LFs consisting of 370×538 sub-aperture images on (central) 7×7 angular ($\boldsymbol{\nu}$-) grid. The locations of the seven sensors were chosen to focus at equally spaced disparities in the interval $[-1, 0.3]$. We reconstructed LFs consisting of 185×269 RGB sub-aperture images on the 7×7 $\boldsymbol{\nu}$-grid.

3.2. Training setup

We used the Adam optimizer [20] to train all the NNs compared in the paper. We set the default learning rate as $3{\times}10^{-4}$; for training the direct regression CNN, we used 5×10^{-4}. In training the all-in-focus synthesis NN in Section 2.1, we used a batch size of 2 and 40 epochs. We used learning rate scheduling to stabilize the training: the learning rate decays by 0.5 at epochs 3, 6, 10, and 20. For joint training of depth estimation and refining NNs, we used a batch size of 1 and 50 epochs. We chose the regularization parameters in (2) as $\lambda_{\mathrm{c}} = 0.005$ and $\lambda_{\mathrm{tv}} = 0.01$.

3.3. MBIR setup

For 4D EP regularizer, we used the hyperbola penalty function, selected the regularization and hyperpola penalty parameter as 1.6×10^5 and 0.38, respectively, and used conjugate gradient descent method with 30 iterations. We reconstructed each color channel of the LF independently.

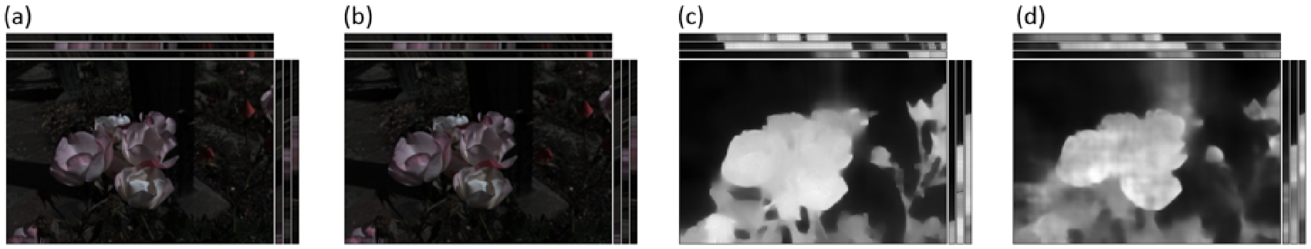

Fig. 2. Sub-aperture images and epipolar slices of the reconstructed LF and the estimated 4D ray depth. (a) Ground truth LF visualized at the corner view. (b) Reconstructed LF via the proposed method at the corner view (PSNR = 42.23 dB). (c) Estimated center view depth via the proposed method. (d) Estimated center view depth via single image sequential CNN [14].

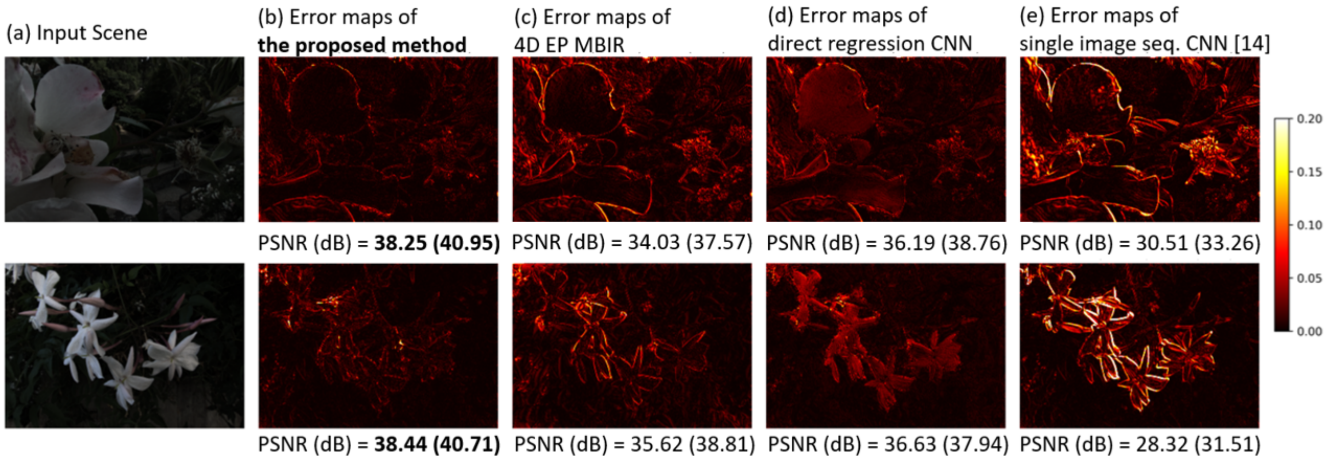

Fig. 3. Error maps of the reconstructed LF sub-aperture view $(u = -1, v = 3)$. The PSNR values shown in parenthesis are calculated from reconstructed LFs.

4. RESULTS

Fig. 2(a-c) shows an example of reconstructed LF and intermediate estimated depth from the proposed method. The proposed method can reconstruct both ray depth and LF with good quality from a focal stack.

Fig. 2(c-d) shows ray depth estimated by the proposed method using focal stack (c) and by sequential CNN using a single image (d). The proposed method can improve depth estimation. As expected, better depth estimates benefits subsequent LF reconstruction: Table 1 shows that the proposed method achieves a 4.8 dB peak signal-to-noise ratio (PSNR) improvement over the state-of-the-art sequential CNN using a single image [14]. In addition, the proposed method significantly improves LF reconstruction accuracy compared to other LF reconstruction methods using focal stack images: the proposed method achieves 2.85 dB and 1.97 dB PNSR improvements, over the conventional 4D EP MBIR method and direct regression CNN, respectively; see Table 1. Fig. 3 shows sub-aperture view error maps of two test LF for all the methods. The error maps of proposed method shows significantly reduced error.

The second column of Table 1 includes the timing comparison between the proposed method and other methods. In particular, it shows that the proposed method significantly reduces computation time compared to 4D EP MBIR.

Methods	PSNR (dB)	Time (sec.)
Proposed method	**39.76**	4.14 (6.2×10^{-2})
Single image sequential CNN [14]	34.96	3.96 (4.6×10^{-2})
4D EP MBIR	36.91	152 (n/a)
Direct regression CNN	37.79	**0.23** (1.7×10^{-3})

Table 1. Average PSNR of the reconstructed LF and reconstruction time (on CPU/GPU) for 100 test samples. Values in parenthesis are GPU reconstruction times.

5. CONCLUSION

This paper proposed a sequential CNN-based framework that reconstructs LF and estimates ray depth from focal stack images. The proposed method achieves significantly more accurate and/or faster LF reconstruction, compared to the state-of-the-art sequential CNN using a single image [14], conventional 4D EP MBIR from focal stack, and direct regression CNN from focal stack.

Future works include applying the proposed method to the LF imaging system using transparent imaging sensors that can capture focal stack with single exposure [21], and investigating effects of the number of focal planes on LF reconstruction and depth estimation performance.

6. REFERENCES

[1] Zhao Pei, Yanning Zhang, Tao Yang, Xiuwei Zhang, and Yee-Hong Yang, "A novel multi-object detection method in complex scene using synthetic aperture imaging," *Pattern Recognition*, vol. 45, no. 4, pp. 1637–1658, Oct. 2012.

[2] Ralph Gross, Iain Matthews, and Simon Baker, "Appearance-based face recognition and light-fields," *IEEE Trans. Pattern Anal. Mach. Intell*, vol. 26, no. 4, pp. 449–465, Feb. 2004.

[3] Ramachandra Raghavendra, Bian Yang, Kiran B Raja, and Christoph Busch, "A new perspectiveface recognition with light-field camera," in *Proc. IEEE Int. Conf. on Biometrics (ICB)*, June 2013, pp. 1–8.

[4] Kiran B Raja, Ramachandra Raghavendra, Faouzi Alaya Cheikh, Bian Yang, and Christoph Busch, "Robust iris recognition using light-field camera," in *Proc. IEEE Colour and Visual Computing Symposium (CVCS)*, Sept. 2013, pp. 1–6.

[5] Anat Levin and Fredo Durand, "Linear view synthesis using a dimensionality gap light field prior," in *Proc. IEEE CVPR*, June 2010, pp. 1831–1838.

[6] Antoine Mousnier, Elif Vural, and Christine Guillemot, "Partial light field tomographic reconstruction from a fixed-camera focal stack," *arXiv preprint arXiv:1503.01903*, May 2015.

[7] Chang Liu, Jun Qiu, and Ming Jiang, "Light field reconstruction from focal stack based on Landweber iterative scheme," in *Proc. Imaging and Applied Optics*, June 2017, pp. MM2C–3.

[8] Cameron J Blocker, Il Yong Chun, and Jeffrey A. Fessler, "Low-rank plus sparse tensor models for light-field reconstruction from focal stack data," in *Proc. IEEE Image, Video, and Multidim. Signal Process. Workshop (IVMSP)*, June 2018, pp. 1–5.

[9] Karen Simonyan and Andrew Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. ICLR*, May 2015.

[10] Il Yong Chun, Xuehang Zheng, Yong Long, and Jeffrey A. Fessler, "BCD-Net for low- dose CT reconstruction: Acceleration, convergence, and generalization," in *Proc. Med. Image Computing and Computer Assist. Interven. (MICCAI)* (to appear), Shenzhen, China, Oct. 2019.

[11] Il Yong Chun and Jeffery A. Fessler, "Deep BCD-Net using identical encoding-decoding CNN structures for iterative image recovery," in *Proc. IEEE Image, Video, and Multidim. Signal Process. Workshop (IVMSP)*, June 2018, pp. 1–5.

[12] Hongki Lim, Il Yong Chun, Yuni K. Dewaraja, and Jeffrey A. Fessler, "Improved low-count quantitative PET reconstruction with a variational neural network," *arXiv preprint arXiv:1906.02327*, May 2019.

[13] Nima Khademi Kalantari, Ting-Chun Wang, and Ravi Ramamoorthi, "Learning-based view synthesis for light field cameras," *ACM Trans. on Graphics (TOG)*, vol. 35, no. 6, pp. 193, Nov. 2016.

[14] Pratul P Srinivasan, Tongzhou Wang, Ashwin Sreelal, Ravi Ramamoorthi, and Ren Ng, "Learning to synthesize a 4D RGBD light field from a single image," in *Proc. IEEE ICCV*, Oct. 2017, pp. 2243–2251.

[15] Anil Kumar Vadathya, Sharath Girish, and Kaushik Mitra, "A unified learning based framework for light field reconstruction from coded projections," *arXiv preprint arXiv:1812.10532*, 2018.

[16] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Med. Image Computing and Computer Assist. Interven. (MICCAI)*, Shenzhen, China, Oct. 2019, pp. 234–241.

[17] Fisher Yu and Vladlen Koltun, "Multi-scale context aggregation by dilated convolutions," in *Proc. ICLR*, May 2016.

[18] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, "Deep residual learning for image recognition," in *Proc. IEEE CVPR*, June 2016, pp. 770–778.

[19] Il Yong Chun, Zhengyu Huang, Hongki Lim, and Jeffrey A Fessler, "Momentum-Net: Fast and convergent iterative neural network for inverse problems," *arXiv preprint arXiv:1907.11818*, Jul. 2019.

[20] Diederik P Kingma and Jimmy Lei Ba, "Adam: A method for stochastic optimization," in *Proc. ICLR*, May 2015, pp. 1–15.

[21] Dehui Zhang, Zhen Xu, Zhengyu Huang, Audrey Rose Gutierrez, Il Yong Chun, Cameron J Blocker, Gong Cheng, Zhe Liu, Jeffrey A. Fessler, Zhaohui Zhong, et al., "Graphene-based transparent photodetector array for multiplane imaging," in *CLEO: Science and Innovations*. Optical Society of America, 2019, pp. SM4J–2.