




Lensless Image Reconstruction with an Untrained Neural Network

Abeer Banerjee^{1,2} , Himanshu Kumar^{1,2}, Sumeet Saurav^{1,2},
and Sanjay Singh^{1,2}

¹ Academy of Scientific and Innovative Research (AcSIR), CSIR-Human Resource Development Centre, (CSIR-HRDC) Campus, Postal Staff College Area, Sector 19, Kamla Nehru Nagar, Ghaziabad 201 002, Uttar Pradesh, India
{abeer.ceeri20a,himanshu.ceeri20a}@acsir.res.in

² CSIR-Central Electronics Engineering Research Institute (CSIR-CEERI), Pilani 333 031, Rajasthan, India
{sumeet,sanjay}@ceeri.res.in
<https://www.acsir.res.in>, <https://www.ceeri.res.in>

Abstract. Lensless image reconstruction is an ill-posed inverse problem in computational imaging, having several applications in machine vision. Existing approaches rely on large datasets for learning to perform deconvolution and are often specific to the point spread function of a particular lensless imager. Generating pairs of lensless images and their corresponding ground truths requires a specialized laboratory setup, thus making the dataset collection procedure challenging. We propose a reconstruction method using untrained neural networks that relies on the underlying physics of lensless image generation. We use an encoder-decoder network for reconstructing the lensless image for a known PSF. The same network can predict the PSF when supplied with a single example of input and ground-truth pair, thus acting as a one-time calibration step for any lensless imager. We used a physics-guided consistency loss function to optimize our model to perform reconstruction and PSF estimation. Our model generates accurate non-blind reconstructions with a PSNR of 24.55 dB.

Keywords: Lensless image reconstruction · Untrained neural networks · Computational imaging

1 Introduction

Cameras have evolved into an important part of human life, acting as external sensors for observing the physical environment. Cameras of with various functional attributes are used for several purposes, ranging from macro photography to astronomical photography. The recent advancements in technologies like computer vision-based wearables, augmented reality, and microrobotics were essential factor that demanded the miniaturizing of imaging systems. The volume and size of cameras have decreased over time for various applications, but the

dependence on lenses prevents further reductions in size because the conventional lens-based optical elements contribute to more than 90% of the volume of the imager. Computational imaging research is attempting to eliminate the need for lenses in the camera arrangement to drastically reduce camera size [1–4]. Removing lenses reduces the size of the camera nearly to the flat form-factor, and various sensor shapes are possible to use.

The lensless imaging setup captures the object information by multiplexing light rays onto the sensors, and the image is then reconstructed by solving an inverse problem. Different optical elements are used for multiplexing, such as diffusers, coded apertures, or diffraction gratings. Existing research in lensless image reconstruction work has produced results comparable to lensed images utilizing two different approaches: optimization-based and learning-based [5]. Lensless cameras have potential use-cases in 3D microscopy [6, 7], monocular depth estimation, and other such fields [8]. They are being employed for privacy-protecting applications [9, 10] because, without the knowledge of the point spread function, the mother image can not be robustly reconstructed with the currently available methods.

Image reconstruction using a lensless setup can be understood as the inverse computational imaging problem. Traditionally, domain-specific recovery algorithms have been utilized to create hand-crafted mathematical models that draw conclusions from their understanding of the basic forward model related to the measurement. These techniques often do not rely on a dataset to learn the mapping, and because the problem is ill-posed, the models frequently show poor discriminative performance [11]. On the other side, deep learning-based methods provide a breakthrough compared to the hand-crafted methods to solve the inverse computational imaging problem [12]. The most recent advancement in Generative Adversarial Networks (GANs) has demonstrated its capability to recreate high-resolution pictures with lesser information from the sample data. With hand-crafted methods, it was impossible to achieve this high compression ratio. However, the success of the deep-learning-based approaches majorly rely on large labelled datasets, and medical imaging and microscopic imaging are a few of the applications in which the enriched labelled data is not available. In contrast to the earlier training-based deep learning approaches, untrained neural networks are able to estimate the faithful reconstruction using the corrupted measurement as an input, without having any prior exposure to the ground truth.

Various optimization algorithms are used to solve inverse problems to recover the original image. The alternating direction method of multipliers (ADMM) [13], regularised L1 [14], or total variation regularisation [15] are common variations that have been modified for lensless imaging [1, 2]. The main advantage of this method is that they are data-agnostic, but there is a trade-off between computational complexity and reconstruction quality. The deep-learning based

ResNet architecture is utilized with raw sensor measurements to create the image [16]. However, the reconstruction of the image is only possible if the acquired image was taken up close since this makes it easier to project high-intensity features onto the sensor. In continuation, deep learning for lensless optics has been coupled with spatial light modulators, scattering media [17], and glass diffusers [18] to produce reconstructions. With the use of lensless images and object-detectors based on Convolutional Neural Networks, FlatCam [19] was able to recognize faces using deep learning. Khan et al. [20] proposed that Flatcam reconstruction was performed using GANs without the need for a point-spread PSF during testing. Another proposed method optimizes unrolled ADMM using a learnable parameter that could be integrated with U-Net and tested on diffuser images [21]. The existing work either relies on the large labelled dataset to perform the reconstruction task or the iterative optimization method which is computationally complex. In this paper, we reconstruct lensless images with untrained neural networks. No training data is used in the reconstruction process that is guided by a physics-informed consistency loss. We verified our approach with images captured using multiple random-diffusers, each with a unique random point spread function. We performed a detailed performance evaluation of our method against the traditional optimization-based and deep-learning-based methods with evaluation metrics like Peak Signal-to-Noise Ratio (PSNR) and Structural SIMilarity index (SSIM). We have also provided visual comparison results that indicate our method was able to outperform the existing methods in the majority of test cases.

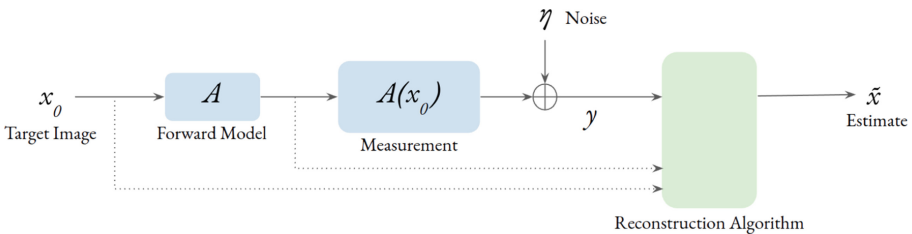


Fig. 1. A typical pipeline for solving inverse problems in computational imaging. The forward model A generates the measurement. The reconstruction algorithm could require priors about the target image, the forward measurement procedure, etc. to reconstruct the estimate \tilde{x} .

2 Theoretical Basis

Lensless image reconstruction is a classic example of an inverse problem in computational imaging. Typically, in every inverse problem, the main task is the reconstruction of a signal for the available observed measurement. In our case,

the observed measurement is the lensless image captured using the bare image sensor with a random diffuser protecting it. A general approach toward solving an inverse problem is to formulate the forward problem as:

$$y = A(x_0) + \eta, \quad (1)$$

where y is the measurement obtained via a forward operation A that signifies the physical measurement process that acts on the input x_0 , and η represents the noise process. In mathematical terms, the inverse problem refers to the faithful reconstruction of the target image x_0 , given the available measurement y . There are a wide variety of inverse problems such as denoising, super-resolution, phase-retrieval, inpainting, and deconvolution that are fundamentally differentiated by the forward operator used for generating the measurement. In this paper, we concern ourselves with lensless image reconstruction, which is essentially a deconvolution problem. The forward operator for a deconvolution problem is formulated as:

$$A(x) = k * x, \quad (2)$$

where k is called the point spread function (PSF), and $*$ denotes the convolution operation. Reconstruction algorithms that utilize the prior knowledge of the PSF that characterizes the whole lensless image formation process, are called non-blind deconvolution techniques. Algorithms that do not require explicit knowledge about the PSF are called blind deconvolution techniques, but they are ill-posed and can result in multi-modal solutions.

Deep Learning based approaches are being extensively used for solving inverse problems for their ability to leverage large datasets for learning a mapping from y to x . Generative adversarial training of U-Nets has resulted in excellent lensless image reconstruction models [22, 23], especially for the non-blind deconvolution task. However, the task-specificity of these discriminative approaches reduces the generalizability of the model and these approaches are yet not well equipped at handling even subtle changes in the forward measurement process. Furthermore, the generation and recording process of datasets related to inverse problems is often complicated. The huge computational cost involved in the retraining and reconfiguration of learning-based methods subject to the availability of datasets with acceptable quality is major setback that need to be addressed.

Physics-informed neural networks are being used for handling ill-posed inverse problems due to their capability of integrating mathematical physics and data. Incorporating the physics of operation of a particular problem simplifies the task and helps in faster convergence of the model, and also attempts to address the problem of big data requirements [24]. Untrained Neural Networks (UNN) completely solve the problem of training data requirements making them perfect for problems where paired data collection is challenging or cumbersome. An approach for phase imaging using untrained neural networks has been explored by Wang et al. [25]. There, they use an iterative approach for optimizing a physics-enhanced deep neural network to produce the object phase. Our approach is

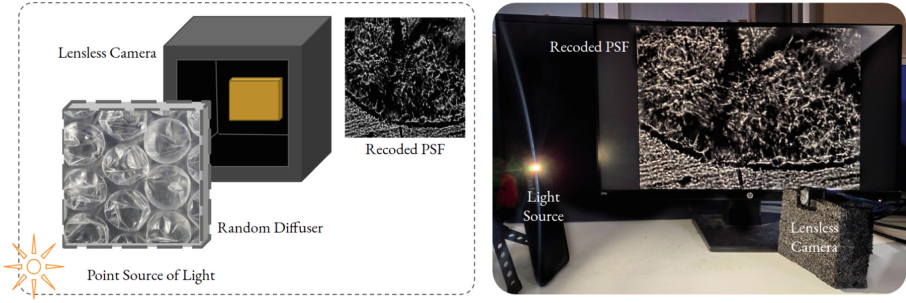


Fig. 2. Experimental setup for capturing the point spread function obtained for a random diffuser. The same setup was used for capturing lensless images corresponding to the lensed images displayed on the monitor.

inspired by their concept of incorporating a physical model into a neural network, which we use to solve the deconvolution problem in lensless computational imaging.

In this paper, we attempt to address the inverse computational imaging problem using a UNN to get reconstruction of the lensless image. We iteratively optimize our model with a physics-informed consistency loss, thereby achieving faster convergence and better performance compared to traditional optimization-based techniques. Our approach leverages random diffusers prepared using inexpensive materials like bubble wraps to capture lensless images making the whole setup compact and easily reproducible.

3 Experimental Setup

Figure 2 illustrates the experimental setup used for capturing the PSF with a random diffuser. We used a bright light source as a point source to illuminate the lensless camera system. We used a random diffuser, i.e., a bubble wrapping plastic sheath, to protect the camera sensor. The resulting PSF recorded using this setup was convolved with a lensed image to generate a lensless image. The lensless image, thus generated, was compared with the lensless image captured using the same setup, and we found that they were very similar. Ideally, they should have matched to each pixel, but practically, the lensless image capturing process introduces a slight amount of noise into the captured image.

4 Methodology

This section discusses the network architecture designed for the reconstruction pipeline. We have not used any training dataset for reconstruction. The images used for explaining the pipeline are obtained from the lensless image test set of nine images captured with the DiffuserCam, provided by Monakhova et al. [21].

4.1 Network Architecture

For a reconstruction task, it is very common to use an encoder-decoder architecture. U-Nets are the most commonly used networks for image reconstruction since they have skip connections that help in modelling identity transformations. However, the lensless images are heavily multiplexed, leading to an image that is incomprehensible to humans and that shares almost no structural attributes with their lensed counterparts. Therefore, the presence of identity connections need not imply a faster convergence, since an identity transformation does not help.

We have used an encoder-decoder framework with the encoder network being popular convolutional architectures like a ResNet or a DenseNet. The CNN architectures available off-the-shelf are truncated to exclude the fully-connected layers, such that the resulting encoder structure is fully convolutional. The structure of the decoder network is mostly fixed with a varying number of input channels according to the feature maps produced by the truncated encoder.

4.2 Pipeline

Figure 3 illustrates the PSF estimation and the reconstruction process. The lensless image reconstruction pipeline follows an untrained iterative optimization that uses a physics-based consistency loss for optimizing the encoder-decoder framework. In the forward path, the lensless image is set as the input to the neural network that produces an intermediate reconstruction y . The output of the neural network is passed through the mathematical process of lensless image generation H , which refers to the convolution of the intermediate reconstruction with the PSF to produce an intermediate lensless image. Theoretically, the neural network achieves perfect reconstruction if the generated intermediate lensless image via process H is the same as the input lensless image. Therefore, we backpropagate the Mean-Squared Error loss between the generated intermediate lensless image \tilde{I} via process H , and the original lensless image x , i.e., the physics-based consistency loss.

The same framework can be repurposed to estimate the PSF, subject to the condition that we already have a pair lensed-lensless pair. This is known as the calibration step that is specific to every lensless camera system. Here, we provide the network with the lensless image and expect the PSF as the output. The model can be forced to predict the PSF by changing the mathematical process H , i.e., by convolving the intermediate output with the available lensed image corresponding to the lensless image. The remaining process of calculating the consistency loss is the same.

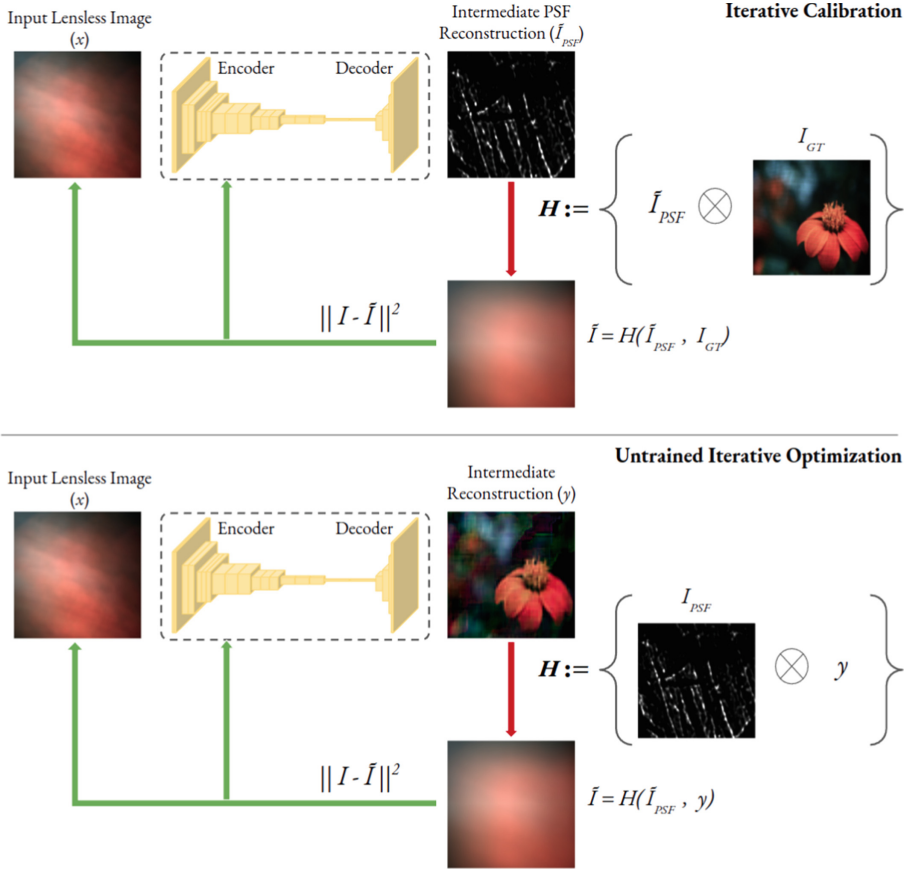


Fig. 3. The complete pipeline for lensless image reconstruction. The top section of the image illustrates the calibration step for PSF estimation using a single pair of lensless and ground truth images. The bottom section illustrates the iterative reconstruction approach. It is to be noted that the same network is being utilized for calibration and reconstruction.

5 Results and Analysis

This section presents the results obtained using the testing set of the Diffuser-Cam dataset [21]. The images were created with a random diffuser, making them suitable for the evaluation of our framework. We have selected five images from the test set to provide our results using two commonly used metrics in the image reconstruction domain, namely, peak signal-to-noise ratio PSNR and structural similarity index SSIM. Table 1 shows the PSNR obtained by our model corresponding to different images, and Table 2 shows the SSIM results.

Table 1. Image-specific PSNR comparison of our method with the existing popular reconstruction methods.

PSNR					
Method	Flower	Face	Butterfly	Flowers	Tokens
ADMM [13]	14.40	11.10	13.02	12.00	13.36
U-Net [26]	16.57	18.51	14.42	17.43	9.83
GAN [22]	19.32	17.48	14.96	17.08	11.21
Ours	24.55	22.95	17.72	16.59	13.09

Table 2. Image-specific SSIM comparison of our method with the existing popular reconstruction methods.

SSIM					
Method	Flower	Face	Butterfly	Flowers	Tokens
ADMM [13]	0.50	0.38	0.44	0.39	0.55
U-Net [26]	0.62	0.61	0.69	0.60	0.43
GAN [22]	0.67	0.73	0.71	0.75	0.42
Ours	0.84	0.80	0.73	0.70	0.57

5.1 Visual Comparison

Figure 4 compares the visual performance of our method against the existing methods. Metrics like PSNR and SSIM are helpful in determining the reconstruction performance, but they often fail to match the human visual perception, as pointed out by Rego et al. [22]. It can be evidently observed that the visual performance of our model is superior to the existing methods in most scenarios. A noteworthy observation was that ADMM performs exceptionally well compared to the other recent models when it comes to the reconstruction of fine features, as seen in the “Token” image. Otherwise, in most cases, it produces dark and grainy reconstructions.

5.2 Ablation Study

We performed an ablation study of the encoder architecture to determine the variation in reconstruction performance. We used different versions of the DenseNet and the ResNet architectures to obtain truncated encoders corresponding to each architecture. These encoders extracted the features from the lensless image, which were decoded to form the reconstructed image. In the increasing order of parameter count, we used DenseNet-121, DenseNet-201, ResNet-50, and ResNet-101 for the reconstruction task. The results are displayed in Fig. 5.

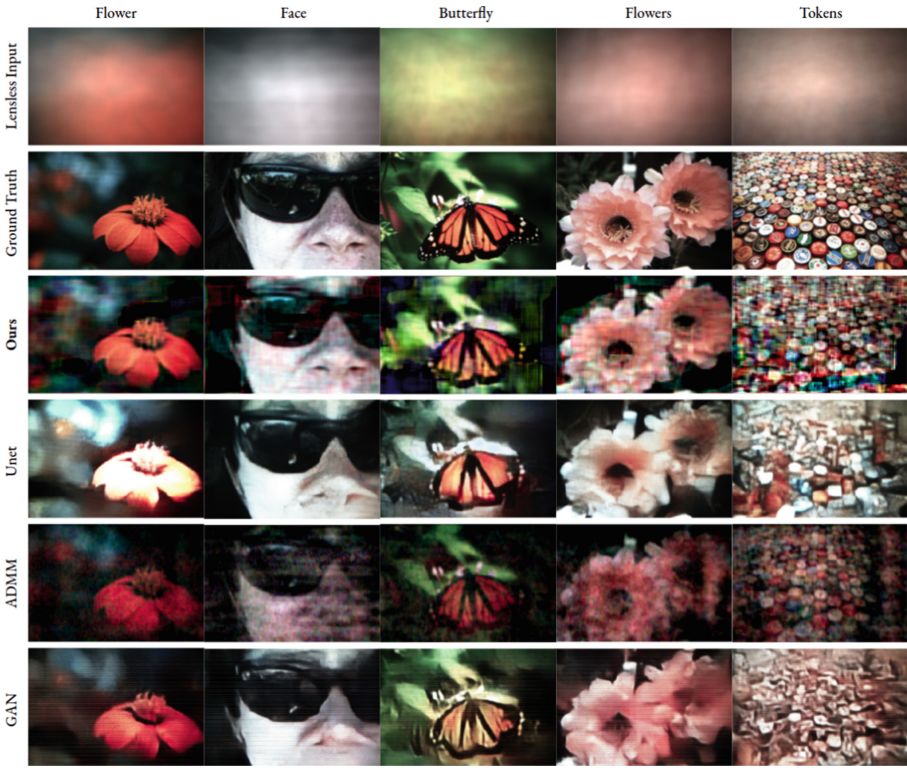


Fig. 4. A visual comparative study of our approach against the existing popular reconstruction approaches.

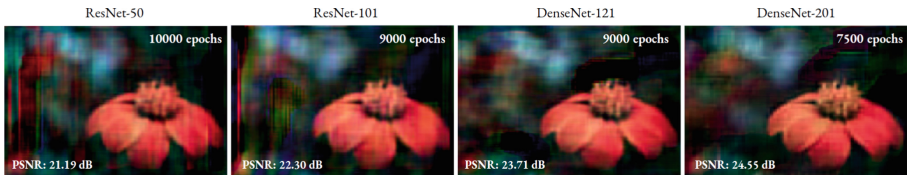


Fig. 5. Ablation study of the encoder backbone used for reconstructing the *Flower* image.

The results obtained by DenseNet-201 appear to be of the best visual quality, but since there is a trade-off between the model performance and the parameter size, DenseNet-121 should be declared the best performer since it achieves a reconstruction performance comparable to the DenseNet-201 with drastically reduced parameter size. The number of epochs for which the iterative optimization had to be performed is plotted on the top right section of each image in Fig. 5. The convergence was significantly faster for the DenseNet framework compared to the ResNet framework, which might be because dense connections strengthen feature propagation, thereby facilitating feature reuse.

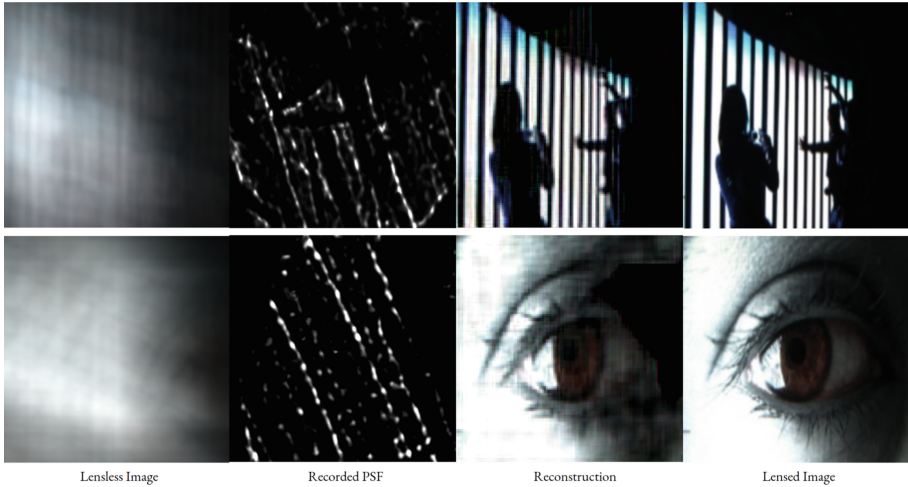


Fig. 6. Reconstructions achieved using PSFs captured by our setup.

5.3 Discussion

Lensless images captured using our own setup took less than 9000 epochs to produce a faithful reconstruction. Figure 6 shows some of the images from the DiffuserCam Training set that were captured using our lensless imager. The PSF was captured using the setup shown in Fig. 2. Our main observation is that the reconstruction of finer features presents inside an image is difficult to reconstruct with the current resolution of PSF convolution. If the PSF convolution is performed at a higher resolution during the physics-based consistency loss calculation, the network might be able to reserve the details, although the optimization time would severely suffer.

6 Conclusion

We have achieved a faithful reconstruction of a lensless image captured using a random diffuser without any training data. We compared the resulting reconstructions with the existing ADMM-based, U-Net-based, and GAN-based approaches and evaluated our performance with metrics like PSNR and SSIM. In almost all cases, our method was able to outperform the existing methods by a significant margin. To support our claims, we present a visual comparison report using the lensless test images provided in the DiffuserCam dataset. The optimization pipeline with the untrained neural networks is a general pipeline that can be repurposed to any inverse imaging task, provided that we have the correct physics-based consistency loss to model the system.

References

1. Antipa, N., et al.: Diffusercam: lensless single-exposure 3d imaging. *Optica* **5**(1), 1–9 (2018)
2. Salman Asif, M., Ayremlou, A., Sankaranarayanan, S.C., Veeraraghavan, A., Baraniuk, R.: FlatCam: Thin, lensless cameras using coded aperture and computation. *IEEE Trans. Comput. Imag.* **3**(3), 384–397 (2017)
3. Tanida, J., et al.: Thin observation module by bound optics (TOMBO): concept and experimental verification. *Appl. Opt.* **40**(11), 1806–1813 (2001)
4. Gill, P.R., Lee, C., Lee, D.-G., Wang, A., Molnar, A.: A microscale camera using direct Fourier-domain scene capture. *Opt. Lett.* **36**(15), 2949–2951 (2011)
5. Boominathan, V., et al.: Lensless imaging: a computational renaissance. *IEEE Signal Process. Mag.* **33**(5), 23–35 (2016)
6. Kuo, G., Antipa, N., Ng, R., Waller, L.: 3D fluorescence microscopy with diffusercam. In: *Imaging and Applied Optics 2018 (3D, AO, AIO, COSI, DH, IS, LACSEA, LS&C, MATH, pcAOP)*, page CM3E.3. Optica Publishing Group (2018)
7. Adams, J.K., et al.: Single-frame 3D fluorescence microscopy with ultraminiature lensless flatscope. *Sci. Adv.* **3**(12) (2017)
8. Zheng, Y., Salman Asif, M.: Joint image and depth estimation with mask-based lensless cameras. In: *2019 IEEE 8th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)* (2019)
9. Wang, Z.W., Vineet, V., Pittaluga, F., Sinha, S.N., Cossairt, O., Kang, S.B.: Privacy-preserving action recognition using coded aperture videos. In: *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 1–10 (2019)
10. Canh, T.N., Nagahara, H.: Deep compressive sensing for visual privacy protection in flatcam imaging. In: *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*, pp. 3978–3986 (2019)
11. Hegde, C.: *Algorithmic Aspects of Inverse Problems Using Generative Models*. IEEE Press (2018)
12. Ongie, G., Jalal, A., Metzler, C.A., Baraniuk, R.G., Dimakis, A.G., Willett, R.: Deep learning techniques for inverse problems in imaging. *IEEE J. Select. Areas Inf. Theory* **1**(1), 39–56 (2020)
13. Boyd, S., Parikh, N., Chu, E., Peleato, B., Eckstein, J.: Distributed optimization and statistical learning via the alternating direction method of multipliers. *Found. Trends Mach. Learn.* **3**(1), 1–122 (2011)
14. Beck, A., Teboulle, M.: A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J. Imaging Sci.* **2**, 183–202 (2009)
15. Rudin, L.I., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. *Phy. D Nonlinear Phenomena* **60**, 259–268 (1992)
16. Sinha, A., Lee, J., Li, S., Barbastathis, G.: Lensless computational imaging through deep learning. *Optica* **4**(9), 1117–1125 (2017)
17. Li, Y., Xue, Y., Tian, L.: Deep speckle correlation: a deep learning approach toward scalable imaging through scattering media. *Optica* **5**(10), 1181–1190 (2018)
18. Li, S., Deng, M., Lee, J., Sinha, A., Barbastathis, G.: Imaging through glass diffusers using densely connected convolutional networks. *Optica* **5**(7), 803–813 (2018)
19. Tan, J., et al.: Face detection and verification using lensless cameras. *IEEE Trans. Comput. Imaging* **5**(2), 180–194 (2018)

20. Khan, S.S., Adarsh, V.R., Boominathan, V., Tan, J., Veeraraghavan, A., Mitra, K.: Towards photorealistic reconstruction of highly multiplexed lensless images. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 7860–7869 (2019)
21. Monakhova, K., Yurtsever, J., Kuo, G., Antipa, N., Yanny, K., Waller, L.: Learned reconstructions for practical mask-based lensless imaging. *Opt. Express* **27**(20), 28075–28090 (2019)
22. Rego, J.D., Kulkarni, K., Jayasuriya, S.: Robust lensless image reconstruction via PSF estimation. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pp. 403–412 (2021)
23. Nelson, S., Menon, R.: Bijective-constrained cycle-consistent deep learning for optics-free imaging and classification. *Optica* **9**(1), 26–31 (2022)
24. George Em Karniadakis, Ioannis G Kevrekidis, Lu Lu, Paris Perdikaris, Sifan Wang, and Liu Yang. *Phys. Inf. Mach. Learn. Nat. Rev. Phys.* **3**(6), 422–440 (2021)
25. Wang, F., et al.: Phase imaging with an untrained neural network. *Light: Sci. Appl.* **9**(1), 1–7 (2020)
26. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28