Unrolled Primal-Dual Networks for Lensless Cameras

Oliver Kingshott,[1] Nick Antipa,[2], Emrah Bostan[3] and Kaan Akşit[1, *]

[1] *University College London, London, United Kingdom*
[2] *University of California San Diego, San Diego, United States of America*
[3] *ams OSRAM, Lausanne, Switzerland*
[*] *k.aksit@ucl.ac.uk*

Abstract: Conventional models for lensless imaging assume that each measurement results from convolving a given scene with a single experimentally measured point-spread function. These models fail to simulate lensless cameras truthfully, as these models do not account for optical aberrations or scenes with depth variations. Our work shows that learning a supervised primal-dual reconstruction method results in image quality matching state of the art in the literature without demanding a large network capacity. We show that embedding learnable forward and adjoint models improves the reconstruction quality of lensless images (+5dB PSNR) compared to works that assume a fixed point-spread function.

1. Introduction

A lensless camera uses a thin mask in place of a conventional lens. Masks can manipulate phase, amplitude, or the entire complex light field of a given scene. Unlike lenses in conventional cameras, these masks can be placed near the imaging sensor, enabling thinner and lighter imaging systems. Additionally, lensless cameras offer the benefits of compressed imaging [1, 2], embedding higher dimensional scene information such as depth from a single capture. To benefit from these qualities, experts typically model lensless cameras as a linear system and recover images computationally by solving the inverse problem.

Pseudo-random phase masks have demonstrated adequate performance for lensless photography [5, 6]. Unfortunately, image reconstruction typically requires computationally expensive and slow iterative reconstruction algorithms (e.g. ADMM [5] and FISTA [7]). To address this, a growing number of works use data-driven Convolutional Neural Networks (CNNs) to improve the speed and quality of lensless image reconstructions [8–10]. A typical CNN with a limited receptive field size fails to accurately model the light transport of the imaging system [11], leading to learned models which fail to reconstruct lensless images accurately and efficiently. Subsequent work using vision transformers have addressed the limited receptive field-size of CNNs, however these require substantial time to train compared to physically informed models [12, 13]. Recent literature proposes neural networks that include a physical model with a large receptive field [6, 14]. These neural networks typically use a single-shot calibration measurement of the Point-Spread Function (PSF) to represent the physical model of the imaging system. However, without the use of precisely engineered masks [6, 15], image formation in lensless cameras cannot be fully expressed by a single PSF model [16]. This model mismatch can lead data-driven regularizers to hallucinate missing features or create overly smooth images. Therefore, the development of models that can correct for model error without increased computational complexity or extensive calibration is of critical importance for the widespread adoption of lensless imaging. Our proposed method replaces ADMM with a learned optimization scheme, improving image quality by reducing model error as opposed to intensive post-processing. The result is a versatile deeply-calibrated lensless imaging architecture that avoids model error in

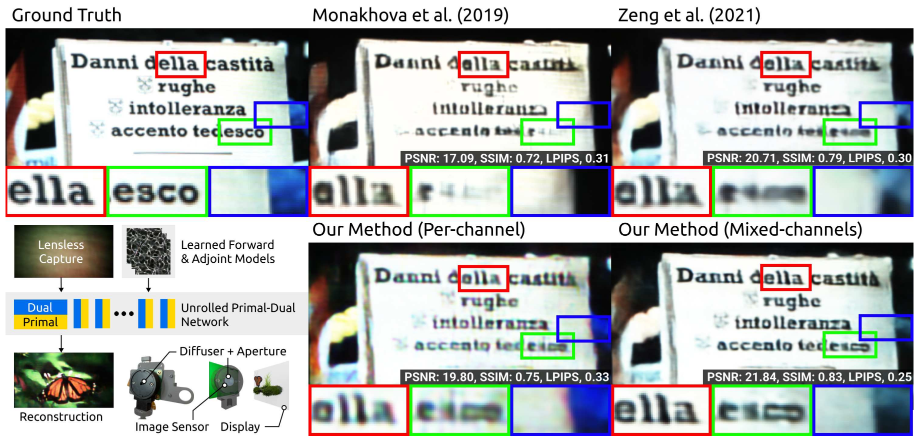

Fig. 1. Comparison of our unrolled primal-dual network with state of the art. Intensive post-processing of lensless images cannot correct the model error, over-smoothing images and removing important features, such as text. We propose to replace classical lensless reconstruction methods with our physically-informed unrolled primal-dual model, where the model includes a series of learned forward and adjoint models (pseudo point-spread functions and their inverse). As a result, our work can produce plausible images and recover additional features while reducing the need for deep post-processing networks such as U-Nets [3] (Source image courtesy MIR Flickr [4]).

the resulting reconstructions. We provide the results of numerous experiments comparing our method against existing image reconstruction algorithms for lensless cameras.

Specifically, our work provides the following contributions:

- Learned primal-dual for lensless imaging. We show for the first time that a modified learned primal-dual optimization framework [17] can recover images from a lensless camera using a pseudo-random phase mask.
- Learned forward-adjoint model. We embed additional linear operators within our learned primal-dual framework. These learned forward-adjoint models are jointly optimized with the rest of our model using the same paired training examples. We show that our extended model provides a significant visual quality enhancement in our image reconstructions. Our method promises reductions up to 50% in reconstruction error while using a fraction of the parameters compared to previous works.
- Lensless camera prototype. We build a proof of concept lensless camera to test further and demonstrate the performance of our model in an actual lensless camera with a pseudo-random mask. We provide an automatic calibration routine that can train our model without the need for an additional camera with a conventional lens.

Limitations When compared to models that use a single calibrated forward model, our method yields an improvement in the quality of lensless image reconstructions. However, a thorough investigation is required to identify explainable links between our learned forward models and physically accurate models in the future. In our experiments with our in-house built camera, we observe a lesser quality in image reconstructions when compared with the state of the art datasets [6, 14]. We believe these originate from the fact that the off-the-shelf diffuser we use does not fully resemble the case that we draw our inspiration from [5]. However, our work significantly improves the image quality both on benchmark datasets [14] and our in-house built camera.

2. Related work

We introduce a novel image reconstruction method for lensless cameras. Here, we provide a brief survey of prior art in lensless cameras, unsupervised lensless image reconstruction methods and learned image reconstruction techniques. Curious readers can read more about lensless cameras through the work by Boominathan et al. [18] and Kavakli et al. [19].

2.1. Lensless cameras

The idea of building cameras without requiring optical lenses has been a long-standing vision for scientists [20] as optical lenses can be bulky, hard to manufacture with great precision, and are typically focused at one plane at a time. The advent of ubiquitous high performance computing and the promise of high dimensional capture has led to a resurgence of interest in lensless cameras. A lensless camera uses a mask as a hardware optical encoder, and is paired with a computational reconstruction algorithm to recover the scene content. Mask based lensless cameras have been demonstrated with coded illumination [21], coded apertures [22, 23], amplitude-only diffraction gratings (e.g., pinhole arrays [24]), photon sieves [25], separable amplitude masks [26], Fresnel Zone plates [27]), phase-only diffraction gratings [5, 28] and metalenses [15]. Additionally, the mask used in a lensless imaging system can also be co-designed with an algorithm that recovers scene information [15]. The depth-varying PSFs of phase mask imaging systems can augment existing 2D imaging sensors with near-field 3D imaging [5]. Alternatively, single-pixel detectors combined with coded illumination patterns can be used for time-based imaging [29, 30].

In our work, we show a lensless camera prototype for experimental validation. Our prototype is similar to the one demonstrated by [5] but differs in implementation details, which we go through in our implementation section.

2.2. Unsupervised Lensless Image Reconstruction Methods

The large spatial extent of the PSFs used in phase-mask based lensless cameras necessitates a cropped convolution model, owing to the limited size of the imaging sensor. By modelling the convolution and the sensor crop as separable sub-problems, the Alternating-Direction Method of Multipliers [5] can be used to recover images using convex optimization. Convex optimization methods such as ADMM are mathematically rigorous, and offer strong guarantees of convergence in contrast to stochastic methods. However, modelling field-varying aberrations is cumbersome process using convex optimization approaches, typically requiring a 10x or greater increase in computational cost [16].

2.3. Learned Lensless Image Reconstruction Methods

The advent of learning-based approaches eases the computational burden of lensless image reconstruction. The work by [14] unrolls five iterations of ADMM and uses a large U-Net [3] to improve perceptual quality. By augmenting a well-known unrolled optimization with learned post-processing, this method clearly separates the role of known physical models and black-box neural networks. However, this approach has a limited ability to correct for model error in the resulting reconstructions, relying on intensive post-processing to achieve plausible reconstructed images. [31] demonstrates a blind deconvolution model for lensless cameras without involving PSF measurements. Blind deconvolution methods are appealing as they aim to eschew the need for laboratory calibration. Our model requires re-training for each phase mask, yielding higher quality lensless reconstructions at the cost of portability. [32] propose a fast learned reconstruction model for lensless cameras. By improving boundary conditions inherent in the sensor crop,

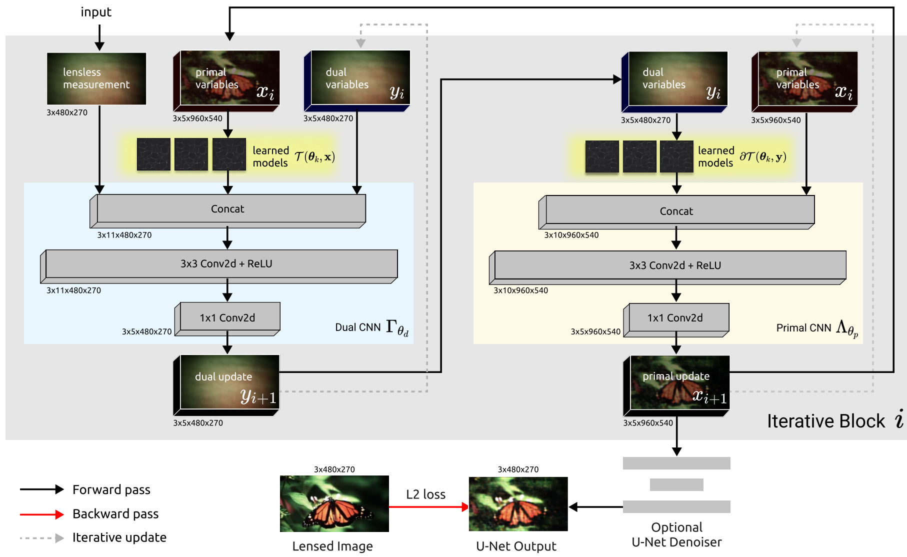

Fig. 2. Unrolled Primal-Dual network architecture for reconstructing lensless images. Our model accepts inputs in the form of a batch of RGB lensless measurements with a predetermined width and height. The blue box illustrates our dual update step, where variables in the measurement domain ($\{y_i, y_{i-1}, b\} \in Y$) are concatenated channel-wise before passing through two convolutional layers parameterized by θ^i_d. The yellow box illustrates our primal update step, where each variable in the measurement domain ($\{x_i, x_{i-1}\}$) is likewise concatenated and convolved with two layers parameterized by θ^i_p. Our forward-adjoint model tensor, θ_k, which is initialized with the value of a PSF measured using a point light source, is also optimized at each epoch. Finally, our trained model reconstructs images from lensless measurements.

they show that they can recover realistic images in a single step without the need for an iterative model. Our work embeds multiple large kernels within an unrolled iterative model to better compensate for optical aberrations. [33] tackles model mismatch caused by imperfect modelling mainly due to spatially-varying PSFs with varying eccentricity. This is acheived by learning residual blocks during each unrolled iteration of ADMM, which are fed into the U-Net denoiser to correct for model error. We show that our method yields accurate intermediate reconstructions by separating the role of the denoising network from the model reconstruction network. Most recently, [34] have proposed pairing multiple Wiener filters with convolutional neural networks to recover accurate images in a lensless microscopy application. However, their method requires a experimental verification for phase-mask based lensless cameras as it targets microscopy.

In conclusion, existing learned methods depend both on accurate PSF calibration and additional training data to develop a suitable image prior. Our method makes better use of supervision by diverting trainable parameters towards improving the underlying physical model of light transport. By learning iterations of an implicit optimization procedure, our method produces accurate intermediate reconstructions that are more consistent with images captured by a lensed camera. To our knowledge, our learned method delivers results that are on-par with the current state of the art in terms of speed and image quality, while offering greater parameter efficiency than previous works.

3. Method

We first introduce the forward model for a phase-mask based imaging system. We then present our proposed lensless image reconstruction model. Finally, we illustrate our deep calibration procedure which captures the necessary dataset for our supervised model-based reconstruction.

3.1. Problem Formulation

We assume that measurements from our imaging system, $\mathbf{b}$, are the result of a linear transformation $\mathbf{A}$ applied to points in the scene $\mathbf{x}$, with some additional noise ϵ:

$$\mathbf{b} = \mathbf{A}\mathbf{x} + \epsilon, \tag{1}$$

where $\mathbf{b}, \mathbf{x}\epsilon$ are vectors.

Each column of $\mathbf{A}$ corresponds to the linear transformation of a single point in the scene, also known as PSFs. Storing PSFs for each point in memory is a demanding task. Rather than storing all PSFs, using an aperture enables the approximation of $\mathbf{A}$ as a cropped convolution with a PSF measured along the optical axis [5]

$$\mathbf{b} = \mathbf{C}(\mathrm{PSF} * \mathbf{x}) + \epsilon \tag{2}$$

. Here, $*$ represents a circular convolution and $\mathbf{C}$ represents a crop down to the size of the imaging sensor. The lateral shifting of the large PSF outside of the bounds of the image sensor necessitates this cropped convolution model. A single experimentally measured PSF is typically used to reconstruct images using the described convolutional forward model [5, 14]. The on-axis PSF is typically measured by shining a point light source along the optical axis of an existing system. Under the assumption that $\mathbf{b}$ is the result of a cropped convolution with an experimentally measured PSF, we recover an estimate of the scene $\mathbf{x}$ by solving a regularized optimization problem:

$$\hat{\mathbf{x}} \leftarrow \arg\min_{\mathbf{x}} \frac{1}{2} \|\mathbf{C}(\mathrm{PSF} * \mathbf{x}) - \mathbf{b}\|_2^2 + \lambda \mathcal{R}(\mathbf{x}), \tag{3}$$

where $\mathcal{R}$ is a regularization function that penalizes unlikely solutions in the presence of noise, with λ controlling the amount of regularization with respect to the data fidelity term.

In this work, we seek to improve the quality of lensless imaging by embedding learnable convolution kernels that are the same size as the PSF within a learned optimization scheme.

3.2. Learning Large Kernels with Physically Informed Networks

Access to paired training examples unlocks a vast landscape of learned reconstruction techniques. Purely data-based architectures, such as U-Nets, typically require large numbers of paired training examples and suffer from poor generalization on unseen data. These limitations can be overcome by incorporating knowledge of physical processes, such as light transport [35], into the neural network architecture. Physically informed networks such as learned primal-dual [17] are highly data-efficient, requiring only a moderate number of training examples, and tend to generalize well to unseen data. With access to paired training examples, but without knowledge of the true linear system $\mathbf{A}$, we propose to train a reconstruction network $\mathcal{G}$ with the goal of minimizing the average mean squared distance to ground truth reconstructions from a lensed camera $\mathbf{x_{gt}}$:

$$\mathcal{L}_{MSE} := \|\mathcal{G}_\theta(\mathbf{b}) - \mathbf{x_{gt}}\|_2^2 \tag{4}$$

In the next section, we explain the design of G_θ. As the focus of our work is to recover the signal encoded in $\mathbf{b}$, we exclusively use mean-squared error as our loss function.

3.2.1. Learned Primal Dual with a Physical Model

We propose a modified learned primal-dual architecture as our learned reconstruction network $\mathcal{G}$ (Equation 4). Figure 2 illustrates how our data and parameters flow through the network. We extend the original work by [17] in three ways. First, we replace the forward operator $\mathcal{T}$ and its adjoint $\partial\mathcal{T}$ with the cropped convolution operation of our lensless camera in Equation (2):

$$
\begin{aligned}
\mathcal{T}(x) &\leftarrow \mathbf{C}(\mathrm{PSF} * x) \\
\partial\mathcal{T}(y) &\leftarrow \mathbf{P}(\mathrm{PSF} \star y),
\end{aligned} \tag{5}
$$

where $\mathbf{P}$ represents zero padding up to twice the size of the imaging sensor, and $\star$ represents circular cross-correlation. $x \in X$ and $y \in Y$ are primal and dual variables respectively, with the former belonging to the domain of reconstructed images X and the latter in the domain of lensless measurements Y.

Second, we allow the PSF to be optimized during training. We initialize $\theta_k \leftarrow \mathrm{PSF}$, allowing the network to modify the physical PSF during training:

$$
\begin{aligned}
\mathcal{T}(x) &\leftarrow \mathbf{C}(\theta_k * x) \\
\partial\mathcal{T}(y) &\leftarrow \mathbf{P}(\theta_k \star y),
\end{aligned} \tag{6}
$$

Finally, we wish to learn multiple kernels to improve our estimate of the true physical system. We choose to learn n convolution kernels, equal to the number of primal and dual variables. Let

$$
\begin{aligned}
\mathbf{x} &= \begin{bmatrix} x^1 & x^2 & \dots & x^n \end{bmatrix} \\
\mathbf{y} &= \begin{bmatrix} y^1 & y^2 & \dots & y^n \end{bmatrix}, \\
\boldsymbol{\theta}_k &= \begin{bmatrix} \theta_k^1 & \theta_k^2 & \dots & \theta_k^n \end{bmatrix},
\end{aligned} \tag{7}
$$

then each primal and dual variable $x^{1 \dots n}, y^{1 \dots n}$ is convolved or cross-correlated with its own learned kernel $\theta_k^{1 \dots n}$

$$
\begin{aligned}
\mathcal{T}(\mathbf{x}) &\leftarrow \mathbf{C}(\boldsymbol{\theta}_k * \mathbf{x}) \\
\partial\mathcal{T}(\mathbf{y}) &\leftarrow \mathbf{P}(\boldsymbol{\theta}_k \star \mathbf{y}).
\end{aligned} \tag{8}
$$

The above modifications result in a variation of the learned primal-dual algorithm with the following update steps:

$$
\begin{aligned}
\mathbf{y}_i &\leftarrow \Gamma_{\theta_d^i}(\mathbf{y}_{i-1}, \mathcal{T}(\mathbf{x}_{i-1}), \mathbf{b}) \\
\mathbf{x}_i &\leftarrow \Lambda_{\theta_p^i}(\mathbf{x}_{i-1}, \partial\mathcal{T}(\mathbf{y}_i)),
\end{aligned} \tag{9}
$$

where $\Gamma_{\theta_d^i}, \Lambda_{\theta_p^i}$ are small convolutional neural networks that are parameterized by each unrolled iteration $i \in 1 \dots 10$. At the end of the unrolled iterations, the variable x_{10}^1 is chosen as our best estimate of $\hat{\mathbf{x}}$.

3.3. Per-channel & Mixed-channel models

To improve the performance of our method against baseline image quality metrics such as PSNR and SSIM, we propose an additional model based on higher dimensional feature maps as opposed to RGB images. Specifically, we replace k learned RGB kernels with $3 \times k$ single channel kernels, allowing for cross-channel communication across feature maps. This results in a model with an increased signal-to-noise performance at the cost of a decrease in subjective color accuracy. We provide a visual comparison of these two models and quantitative metrics in our results section.

4. Implementation

In this section we document the development of our own lensless camera as shown in Figure 1. Additional details are provided in the supplementary material.

Camera Design We use a Raspberry Pi High-Quality camera connected to a Raspberry Pi Zero W. This specific camera features a removable lens housing which we replaced with our own 3D printed design. Following [14], we used a 0.5 degree engineered diffuser as our mask, placed ~10mm away from the image sensor. Our 3D printed housing is also illustrated in Figure 1. Our custom housing ensures that the optical element is placed at the desired distance from the imaging sensor, and contains space for an optional infrared filter.

Data Capture To capture a training and test dataset, we place our camera ~15cm away from a 5.5 inch OLED display. We illuminate a 5x5 square grid of pixels in the center of the display and capture the resulting image to measure the on-axis PSF. We then use FISTA [7] to reconstruct a test image. This test image is used to estimate a homography that warps each ground truth image to match the perspective of the lensless camera. Automated software shows a variety of images from the DIV2K dataset [36], capturing 8000 training images and 1000 test images.

5. Evaluation

We first present the results of comparing our method against two central state-of-the-art work that uses DiffuserCam dataset [14, 33]. We additionally perform ablation studies to determine the contribution from each component in our method on reconstructed image quality. Finally, we verify our method using our hardware prototype.

5.1. DiffuserCam results

We compare our model's results against the work that uses DiffuserCam dataset [14] in Table 1, where the number of parameters used, the size of training and testing examples, processing time, and image quality are considered.

Our results suggest that our proposed method improves the quality of images reconstructed from measurements captured by a lensless camera. This is supported by qualitative results in Figure 3, which appear to reproduce features that are more faithful to the original ground truth images.

5.2. Ablation Studies

Disabling U-Net Denoiser. To further confirm that the quality of our reconstructions has increased as a result of correcting for model error, we measure the quality of intermediate reconstructions without the use of a U-Net for denoising. We show our qualitative results in Figure 4 and quantitative results in Table 1. When our U-Net is disabled, the resulting images are noisy but are faithful to the ground truth images. Our intermediate reconstructions demonstrate that our model-based reconstruction network performs the bulk of the work in producing usable lensless reconstructions.

Effect of learning multiple models We ran an additional study to quantify the effect of decreasing the number of learned models from 5 to 1. We include quantitative results in Table 1 and present reconstructed images from our reduced model in Figure 4. Decreasing the number of learned models from 5 to 1 decreases the resulting image quality after post-processing by ~2dB.

Method	PSFs	U-Net	PSNR	LPIPS	Parameters	Runtime (ms)	Training Examples	Iterations
ADMM	1 RGB (fixed)		11.97	0.60	-	1190	0	100
Le-ADMM	1 RGB (fixed)		11.89	0.57	20	50	100	5
Le-ADMM-U	1 RGB (fixed)	✓	20.46	0.37	10.6M	55	24,000	5
Ours (RGB)	1 RGB (learned)		16.74	0.54	0.4M	74	9,000	10
		✓	21.47	0.43	1.2M	77	9,000	10
Ours (RGB)	5 RGB (learned)		16.91	0.51	2.0M	80	9,000	10
		✓	23.48	0.40	2.7M	88	9,000	10
Ours (Mixed)	15 (learned)		19.00	0.48	2.0M	82	9,000	10
		✓	25.34	0.35	3.8M	84	9,000	10

Table 1. Comparison of our models against previous work by [14]. Our model achieves produces modestly accurate reconstructions quickly without the use of a large U-Net, at the cost of learning additional large kernels θ^k. These kernels occupy the majority of our parameter space. Adding a small U-Net to our models improves reconstruction quality further. Increasing the number of learned kernels improves PSNR by ~2dB when combined with U-Net denoising, with cross-channel denoising adding another ~2dB.

5.3. Prototype results

We additionally compare the results of our learned model using a prototype camera built in the lab. We present sample reconstructions in Figure 5 and provide additional reconstructions in our supplementary material.

6. Discussion

Comparison to classical methods. Our proposed models are end-to-end differentiable. They are trained to learn an unrolled iterative reconstruction algorithm, a physically informed model, and a suitable image prior. While our model appears to produce accurate intermediate reconstructions, it is difficult to discretely map each learned component of the model to a specific component existing classical methods. One line of future work could be to establish whether embedding learnable physical models within a classical variational method can achieve similar results. A forward model that is learned independently of image priors and a chosen reconstruction algorithm could be used to evaluate the data fidelity of reconstructed images against their lensless measurements. While the need for supervision in the form of paired image examples is cumbersome, the accuracy of the recovered images is clearly improved.

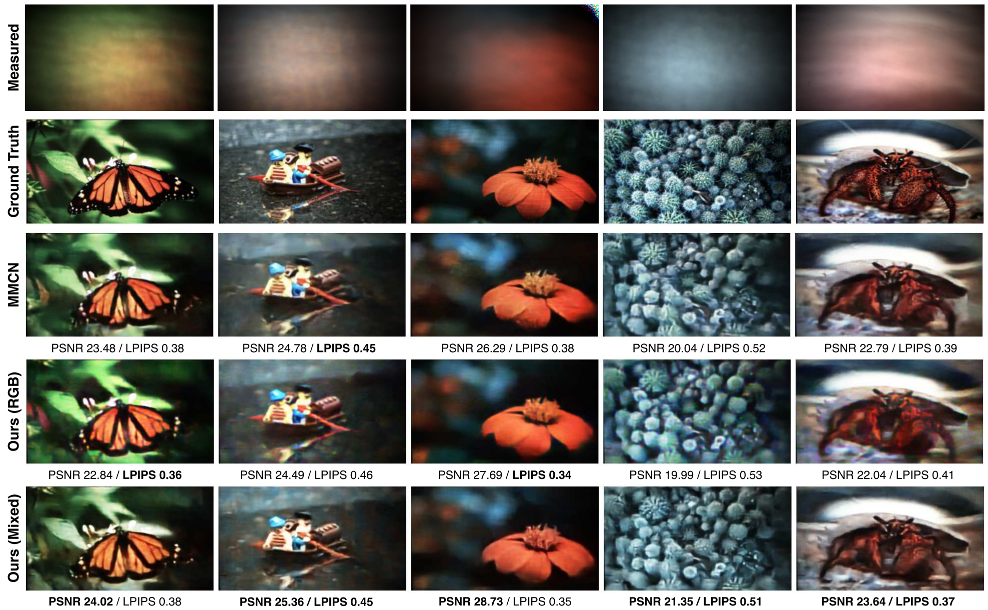

Fig. 3. Comparison of reconstructed test images against the ground truth images. We compare our method against MMCN [33]. MMCN is based on five unrolled iterations of ADMM with additional residual blocks to correct for model error. Our per-channel model (RGB) improves subjective color accuracy while our mixed-channel model (Mixed) recovers higher frequency content. The primary feature of both models is that multiple large kernels are learned to correct for model error.

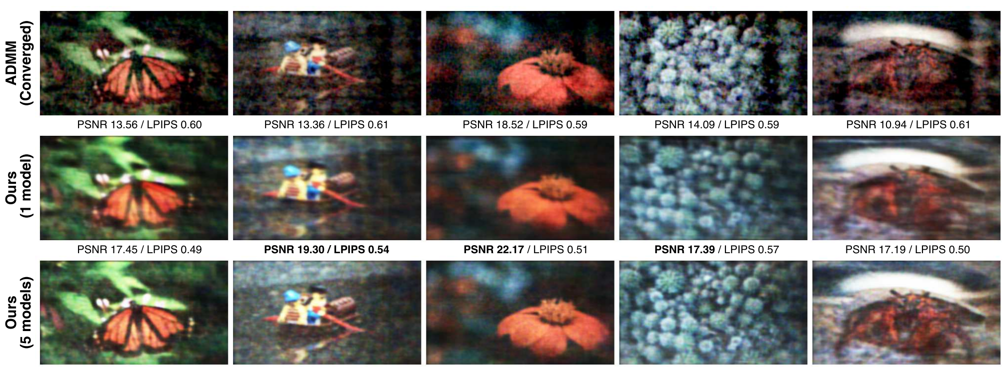

Fig. 4. Comparison of our learned model-based reconstruction networks against unsupervised ADMM (converged) [5]. U-Net denoising was disabled to show that our intermediate reconstructions consist of images that are more faithful to the ground truth data. Learning additional kernels appears to improve accuracy while yielding results faster than classical methods. We reason that our network prioritises consistency with the true physical model, resulting in fewer artifacts.

Comparison to learned methods. When compared to learned methods that use a fixed PSF calibration measurement, our method is able to reconstruct images that more closely resemble images captured by a lensed camera. It is clear that the improved performance of our method is achieved by redistributing model parameters away from deep neural networks and towards the

Fig. 5. Reconstructions from our lensless camera prototype trained using our RGB model. Our optical element consists of a thick holographic diffuser (0.76mm) with bulk scattering, leading to a degradation in image quality.

underlying physical model of light transport in lensless cameras. Additionally, in comparison to methods that use unrolled ADMM iterations such as [14], we find that our method is robust to zero initialization of all model parameters. However, the exact mechanism through which our model improves performance against existing learned methods is unclear. It is possible that our model could be correcting for field-varying aberrations that are not captured by a single on-axis calibration measurement. However, we note that our proposed methods lack any explicit mechanism to apply each learned model to a specific spatial region. Finally, we note that our claim of improved data fidelity can only be measured implicitly by comparing our reconstructions with a lensed camera. In future work, we would like to use measured or simulated field-varying PSFs to design robust models that can explicitly correct for field-varying aberrations without the need for manual calibration.

Color Accuracy. Our two proposed models highlight a potential trade-off between the recovery of high frequency details and color accuracy in our chosen network architecture. Allowing the mixing of color channels appears to increase the frequency content of recovered images. However, our informal subjective opinion is that our per-channel model is able to reproduce color more accurately. We suspect that our per-channel model is vulnerable to color fringing artifacts introduced by the chosen phase masks. Future work could investigate treatment through the use of additional loss functions (such as those proposed by [37]), improved phase mask design [6], or improved network architectures [13].

7. Conclusion

Unconventional camera designs with thin masks in place of conventional lenses offer freedom from the constraints of traditional optics. However, the speed of reconstruction and image quality in mask-based lensless camera designs remains a significant drawback. We argue that neural networks with learnable physical priors for lensless imaging can help to counter this drawback. We show that such hybrid models can provide on-par image reconstruction quality with limited supervision, and without demanding extensive resources in training. We hope that our work can motivate the development of performant and interpretable methods for lensless image reconstruction.

Data and materials availability

All data needed to evaluate the conclusions in the manuscript are provided in the manuscript. Additional data related to this paper may be kindly requested from the author.

Acknowledgement

The authors would like to thank reviewers for their valuable feedback. We thank Laura Waller, Kristina Monakhova, Tianjiao Zeng and Edmund Lam for their support in providing useful insights from their work; Tobias Ritschel for fruitful discussions at the early phases of the project; Koray Kavaklı for his support in hardware prototype related figure and camera homography related software; Tim Weyrich for dedicating GPU resource. Kaan Akşit and Oliver Kingshott relied on the Royal Society's RGS\R2\212229 - Research Grants 2021 Round 2 for building the hardware prototype.

Disclosures

The authors declare no conflicts of interest.

See Supplement 1 for supporting content.

References

1. R. Fergus, A. Torralba, and W. T. Freeman, "Random lens imaging," Tech. rep., MIT (2006).
2. A. Liutkus, D. Martina, S. Popoff, G. Chardon, O. Katz, G. Lerosey, S. Gigan, L. Daudet, and I. Carron, "Imaging with nature: Compressive imaging using a multiply scattering medium," Sci. reports **4**, 1–7 (2014).
3. O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention,* (Springer, 2015), pp. 234–241.
4. M. J. Huiskes and M. S. Lew, "The mir flickr retrieval evaluation," in *Proceedings of the 1st ACM international conference on Multimedia information retrieval,* (2008), pp. 39–43.
5. N. Antipa, G. Kuo, R. Heckel, B. Mildenhall, E. Bostan, R. Ng, and L. Waller, "Diffusercam: lensless single-exposure 3d imaging," Optica **5**, 1 (2018).
6. V. Boominathan, J. K. Adams, J. T. Robinson, and A. Veeraraghavan, "Phlatcam: Designed phase-mask based thin lensless camera," IEEE Trans. on Pattern Anal. Mach. Intell. **42**, 1618–1629 (2020).
7. A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," SIAM J. on Imaging Sci. **2**, 183–202 (2009).
8. A. Sinha, J. Lee, S. Li, and G. Barbastathis, "Lensless computational imaging through deep learning," Optica **4**, 1117–1125 (2017).
9. G. Barbastathis, A. Ozcan, and G. Situ, "On the use of deep learning for computational imaging," Optica **6**, 921–943 (2019).
10. D. Bae, J. Jung, N. Baek, and S. A. Lee, "Lensless imaging with an end-to-end deep neural network," in *2020 IEEE International Conference on Consumer Electronics - Asia (ICCE-Asia),* (2020), pp. 1–5.
11. J. W. Goodman, *Introduction to Fourier optics* (Roberts and Company Publishers, 2005).
12. X. Pan, X. Chen, T. Nakamura, and M. Yamaguchi, "Incoherent reconstruction-free object recognition with mask-based lensless optics and the transformer," Opt. Express **29**, 37962–37978 (2021).
13. X. Pan, X. Chen, S. Takeyama, and M. Yamaguchi, "Image reconstruction with transformer for mask-based lensless imaging," Opt. Lett. **47**, 1843–1846 (2022).
14. K. Monakhova, J. Yurtsever, G. Kuo, N. Antipa, K. Yanny, and L. Waller, "Learned reconstructions for practical mask-based lensless imaging," Opt. Express **27**, 28075 (2019).
15. E. Tseng, S. Colburn, J. Whitehead, L. Huang, S.-H. Baek, A. Majumdar, and F. Heide, "Neural nano-optics for high-quality thin lens imaging," arXiv preprint arXiv:2102.11579 (2021).

16. K. Yanny, N. Antipa, W. Liberti, S. Dehaeck, K. Monakhova, F. L. Liu, K. Shen, R. Ng, and L. Waller, "Miniscope3d: optimized single-shot miniature 3d fluorescence microscopy," Light. Sci. & Appl. **9**, 1–13 (2020).
17. J. Adler and O. Öktem, "Learned primal-dual reconstruction," IEEE transactions on medical imaging **37**, 1322–1332 (2018).
18. V. Boominathan, J. T. Robinson, L. Waller, and A. Veeraraghavan, "Recent advances in lensless imaging," Optica **9**, 1–16 (2022).
19. K. Kavakli, D. R. Walton, N. Antipa, R. Mantiuk, D. Lanman, and K. Akşit, "Optimizing vision and visuals: lectures on cameras, displays and perception," in *ACM SIGGRAPH 2022 Courses,* (2022), pp. 1–66.
20. H. Barker, "Pin-hole or lensless photography," J. Royal Astron. Soc. Can. **14**, 16 (1920).
21. Y. Zheng and M. S. Asif, "Coded illumination for improved lensless imaging," arXiv preprint arXiv:2111.12862 (2021).
22. R. Horisaki, Y. Okamoto, and J. Tanida, "Deeply coded aperture for lensless imaging," Opt. Lett. **45**, 3131–3134 (2020).
23. M. S. Asif, A. Ayremlou, A. Sankaranarayanan, A. Veeraraghavan, and R. G. Baraniuk, "Flatcam: Thin, lensless cameras using coded aperture and computation," IEEE Trans. on Comput. Imaging **3**, 384–397 (2017).
24. V. Anand, S. H. Ng, J. Maksimovic, D. Linklater, T. Katkus, E. P. Ivanova, and S. Juodkazis, "Single shot multispectral multidimensional imaging using chaotic waves," Sci. reports **10**, 1–13 (2020).
25. A. Ö. Yöntem, J. Li, and D. Chu, "Imaging through a projection screen using bi-stable switchable diffusive photon sieves," Opt. express **26**, 10162–10170 (2018).
26. M. J. DeWeert and B. P. Farm, "Lensless coded-aperture imaging with separable doubly-toeplitz masks," Opt. Eng. **54**, 023102 (2015).
27. J. Wu, H. Zhang, W. Zhang, G. Jin, L. Cao, and G. Barbastathis, "Single-shot lensless imaging with fresnel zone aperture and incoherent illumination," Light. Sci. & Appl. **9**, 1–11 (2020).
28. S. Bernet, W. Harm, A. Jesacher, and M. Ritsch-Marte, "Lensless digital holography with diffuse illumination through a pseudo-random phase mask," Opt. express **19**, 25113–25124 (2011).
29. G. Huang, H. Jiang, K. Matthews, and P. Wilford, "Lensless imaging by compressive sensing," in *2013 IEEE International Conference on Image Processing,* (2013), pp. 2101–2105.
30. G. Satat, M. Tancik, and R. Raskar, "Lensless imaging with compressive ultrafast sensing," IEEE Trans. on Comput. Imaging **3**, 398–407 (2017).
31. J. D. Rego, K. Kulkarni, and S. Jayasuriya, "Robust lensless image reconstruction via psf estimation," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision,* (2021), pp. 403–412.
32. S. S. Khan, V. Sundar, V. Boominathan, A. Veeraraghavan, and K. Mitra, "Flatnet: Towards photorealistic scene reconstruction from lensless measurements," IEEE Trans. on Pattern Anal. Mach. Intell. p. 1–1 (2020).
33. T. Zeng and E. Y. Lam, "Robust reconstruction with deep learning to handle model mismatch in lensless imaging," IEEE Trans. on Comput. Imaging **7**, 1080–1092 (2021).
34. K. Yanny, K. Monakhova, R. W. Shuai, and L. Waller, "Deep learning for fast spatially varying deconvolution," Optica **9**, 96–99 (2022).
35. K. Kavaklı, H. Urey, and K. Akşit, "Learned holographic light transport," Appl. Opt. **61**, B50–B55 (2022).
36. R. Timofte, S. Gu, J. Wu, and L. Van Gool, "Ntire 2018 challenge on single image super-resolution: Methods and results," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops,* (2018).
37. F. Heide, M. Rouf, M. B. Hullin, B. Labitzke, W. Heidrich, and A. Kolb, "High-quality computational imaging through simple lenses," ACM Trans. on Graph. (TOG) **32**, 1–14 (2013).