Deep Camera Obscura: An Image Restoration Pipeline for Lensless Pinhole Photography

Joshua D. Rego\textsuperscript{1} \quad Huaijin Chen\textsuperscript{2} \quad Shuai Li\textsuperscript{2} \quad Jinwei Gu\textsuperscript{2} \quad Suren Jayasuriya\textsuperscript{1}

\textsuperscript{1}Arizona State University, Tempe, AZ \quad \textsuperscript{2}SenseBrain Technology, San Jose, CA

\textsuperscript{1}\{jdrego, sjayasur\}@asu.edu \quad \textsuperscript{2}\{chenhuaijin, shuai.li, gujinwei\}@sensebrain.ai

Abstract

The lensless pinhole camera is perhaps the earliest and simplest form of an imaging system using only a pinhole-sized aperture in place of a lens. They can capture an infinite depth-of-field and offer greater freedom from optical distortion over their lens-based counterparts. However, the inherent limitations of a pinhole system result in lower sharpness from blur caused by optical diffraction and higher noise levels due to low light throughput of the small aperture, requiring very long exposure times to capture well-exposed images. In this paper, we explore an image restoration pipeline using deep learning and domain knowledge of the pinhole system to enhance the pinhole image quality through a joint denoise and deblur approach. Our approach allows for more practical exposure times for hand-held photography and provides higher image quality, making it more suitable for daily photography compared to other lensless cameras while keeping size and cost low. This opens up the potential of pinhole cameras to be used in smaller devices, such as smartphones.

1. Introduction

Cameras today have come a long way in enabling high-quality photography capabilities to be easily accessible. Smartphones now incorporate multiple cameras with sensors that can exceed 50MP and enable pro-level features such as HDR mode, panorama stitching, and augmented reality. While these advancements have greatly improved the creativity and quality of everyday photography, lens-based cameras can have their own limitations. Traditionally, smartphone image quality has been largely limited by the...
Figure 2: **System Architecture:** Due to diffraction and low light throughput, pinhole camera images observe two major artifacts: optical blur and sensor noise, for which we propose a jointly optimized denoise and deblur framework to tackle. The proposed system turns blur into an advantage for better denoising. Knowing a circular pinhole automatically performs ideal low-pass filtering (LPF) in optics (see sec. 3), the denoise module first performs optics-aware LPF with an ideal LPF that matches the pinhole’s diffraction limit, since any signal with a frequency above the limit is due to noise. A denoise network then further cleans mid- and low-frequency noise, given deblur or deconvolution process is commonly prone to noise. Finally, a GAN-based deblur network is used to recover high-frequency information lost in the pinhole imaging process to form the final output. Both modules are jointly trained with data-driven loss and physical consistency to eliminate GAN-style artifacts.

There are additional advantages of pinhole cameras, over their lens-based counterparts, that make them an appealing option for consumer photography applications. Pinhole cameras capture an infinite depth-of-field, useful for applications that require all-in-focus images and offer greater freedom from optical distortion due to lenses [66]. The smaller size advantage of lensless pinhole cameras may also allow implementing larger sensors for smartphone photography, such as micro 4/3, APS-C, or full-frame sensors. These advantages yield a potential direction for pinhole cameras to be used for smartphone photography.

However, there are several limiting factors to realizing pinhole cameras for conventional smartphones. Pinhole cameras have very low light throughput (eg. f/200) which introduces higher levels of noise. To appropriately expose an image, long shutter speeds are required which can introduce motion-blur artifacts for moving scenes. Further, the sharpness of the captured image is dictated by the level of optical blur caused by the diffraction of light through the pinhole. While both these factors have limited the deployment of practical pinhole cameras in the past, recent advances in image processing technology including higher sensitivity sensors for low-light imaging, better denoising and deblurring algorithms, and data-driven/machine learning methods present an opportunity to revisit the pinhole camera.
In this paper, we aim to establish the pinhole camera as a compelling candidate for lensless photography, particularly for size-limited applications. We propose a full system pipeline for improving the quality of pinhole photography at short exposures. We leverage both imaging physics and data-driven networks to build this pipeline. A big obstacle for any deblur (or deconvolution operation in general) is noise which affects pinhole cameras due to their small aperture size. However, the aperture function of the pinhole is a circle, which results in an optical point spread function (PSF) of an ideal low pass filter (LPF) of sinc on the imaging plane [24]. Thus any signal higher than the cutoff frequency for this LPF must be noise, and we can simply apply a matching LPF in the digital domain to remove any high-frequency noise without affecting the low-frequency signal.

We leverage this information about a pinhole’s optical point spread function (PSF) to design an imaging system consisting of joint denoising and deblurring for low-light pinhole images. This pipeline is trained on synthetic data but shown to effectively generalize to real-world data, and uses known optical priors including the point spread function (PSF) of the pinhole to help constrain the network’s loss functions. We define practical pinhole photography to be images captured at a usable exposure time of 1/30s, while allowing freedom of a higher ISO which can later be denoised through the pipeline. Our specific contributions include the following:

- An end-to-end imaging pipeline for practical pinhole photography with joint denoising and deblurring for low-light capture
- Reblur losses with a pinhole point spread function (PSF) for improved performance
- A real-world pinhole image dataset with measured HDR PSF suitable for generating synthetic data
- Ablation studies for low-exposure images, including the effects of ISO and exposure time.

We validate our pipeline by comparing against both conventional denoising/deblurring algorithms, traditional optimization-based lensless reconstruction algorithms, and end-to-end lensless camera solutions such as diffusion or coded mask cameras [4, 5]. Our method can handle high-resolution captures of 20 megapixels which is much higher than demonstrated by other lensless camera systems. Our main argument is that practical 2D lensless photography can be achieved with the relatively simpler pinhole camera as opposed to multiplexed imaging methods. We hope this work renews interest in practical pinhole photography in general.

2. Related Work

Pinhole cameras. Pinhole cameras have been around since antiquity with the first recorded mentions by the Chinese philosopher Mozi, Aristotle, Euclid, and Ibn al-Haytham among others [26]. During the Renaissance, Giambattista della Porta, Leonardo da Vinci, and Johannes Kepler can be credited with popularizing the camera obscura [19, 16]. However, with the invention of lenses, film, and digital sensors, pinhole cameras largely fell out of common use due to their low-light and optical blur quality.

There has been much research on pinhole optics, particularly in characterizing effects such as diffraction on image quality [66]. The optimum size for the pinhole for mitigating optical blur has been determined for large pinholes using ray optics [27] and small pinholes using Fraunhofer diffraction [35]. Much work has gone into examining this tradeoff between ray and diffraction optics for pinhole size [52]. There has also been extensive study into the transfer function of pinhole cameras [55] and the preservation of spatial frequencies. In this paper, we show that such traditional optical tradeoffs can be overcome with the use of computational imaging techniques.

Pinhole cameras have been used extensively in scientific imaging applications [47, 20, 30]. Computer vision owes a large debt to the pinhole camera, particularly in the camera model that underpins many geometric vision algorithms [54]. However, papers which leverage actual pinhole cameras, as opposed to the model, for computer vision have been sparse in the literature. Some recent inspiring work has included accidental pinhole and pinspeck cameras [60, 61] showing that simple image processing could be used for apertures naturally occurring in an environment.

Coded Lensless Cameras. Instead of single pinhole cameras, a bulk of research in the computational imaging and photography community has multi-aperture or multiplexed lensless imaging systems. These systems typically allow more light throughput but typically at the expense of worse reconstructed 2D image quality, as we will demonstrate later in our experimental comparison to some systems. However, these systems do have the express advantage of capturing 3D information through their optical multiplexing which is useful for certain applications.

Recently, lensless cameras with an optical element above the sensor have been introduced where these optical elements are thin and scalable to a small size [58, 6, 3, 22]. FlatCam [6] uses a coded amplitude mask to multiplex light from the scene onto the sensor and then reconstructs the image in post-processing. Since amplitude masks lose some light efficiency, newer designs have featured phase masks [8, 62] for improved performance. Alternatively, diffraction gratings [22, 23, 29] and Fresnel plates [36] have
been used to achieve small form factors. Finally, diffusion layers that scatter light onto the sensor have been deployed for lensless imaging systems [3].

To reconstruct lensless images, optimization algorithms are typically deployed to solve the inverse problem. These typically include a variation on either alternating direction method of multipliers (ADMM) [9] or some regularized $\ell_1$ [7] or total variation regularization [49], which are adapted for lensless imaging [6, 3]. Recently deep learning has shown superior performance at lensless image reconstruction and other vision tasks [53, 40, 41]. We compare against the end-to-end lensless camera systems for FlatNet [32] and DiffuserCam [44] in this paper.

Image Deblurring and Denoising. Image denoising and deblurring are traditional low-level vision tasks with a rich history of research. For denoising, we point the reader to [45] for a thorough survey of traditional algorithms. Block-matching-based approaches such as BM3D [17] and non-local mean [10, 39] are top-performing traditional methods widely used nowadays, but are recently exceeded by modern deep learning networks [68, 70, 69]. Leveraging these advances in image denoising research, recent work has changed the photography landscape with successful applications in low-light and burst photography [13, 28]. While many denoisers, including learning-based ones, consider a sensor’s noise formation model to improve the results on real-world images [25, 1], they rarely leverage domain knowledge of the optics to denoise better. In contrast, we utilize the optics’ frequency response to further improve our learning-based denoiser. While we only focus on pinhole imaging here, this idea can be easily extended to any diffraction-limited system, an increasingly important scenario as pixel size shrinks for modern sensors.

For deblurring, it is commonly tackled through either blind or non-blind image deconvolution [11, 36]. For imaging systems, motion blur and defocus blur are the two large categories of deblurring research. Techniques to handle motion blur in photography include traditional [46, 15] and neural methods [12, 63, 21] including the DeblurGAN architectures [37, 38] with high-quality performance. Defocus deblurring research has focused primarily on deconvolution problems with applications in microscopy [51], dual pixels [2], and coded aperture systems [72, 71]. In addition, reblur loss for motion blur [43, 14] is a technique where the static image is reblurred to match the input data. Reblur losses have also been used for lensless cameras [44, 48]. In our paper, we also deploy a reblur loss based on our measured optical point spread function.

Joint denoise/deblur methods have also been achieved for traditional lensed camera photography, particularly to tackle low-light noise and motion blur through burst capture [42]. However, we note that there is still relatively less literature on joint denoising and deblurring. This is probably due to the competing natures of the two tasks: one is aimed at reducing high-frequency content (noise) in the image, and the other aims to restore high-frequency content (spatial details) lost due to optical blur. We also observed this phenomenon that denoising and deblurring algorithms can sometimes undo one another’s effects, and thus it takes careful tuning and understanding of the frequency content of the signal to make these pipelines work together. That is indeed what we show in this paper by leveraging the frequency cutoff of the optical PSF of pinhole cameras to help deblurring perform better in practice.

3. Approach

Our proposed approach involves four critical components: (1) accurate optical point spread function (PSF) capture and modeling, (2) a denoising architecture to improve the low-light captures, (3) high-quality deblurring to restore image details lost due to optical blur, and (4) a reblur loss using the PSF to help jointly train the architecture. A summary of our pipeline can be visualized in Figure 2.

3.1. Forward Image Model with Optical Point Spread Function

The optical point spread function (PSF) represents the impulse response of our pinhole imaging system. It determines how a single infinitely small point source in the scene is received at the image sensor. Due to the diffraction effect, the image for a point source is a spatially spread out bright disk surround by concentric rings of light (an Airy pattern). For a wavelength of light $\lambda$, an imaging system with focal length $f$ and aperture of radius $R$ will produce at the sensor plane an Airy disk width $w$ shown by,

$$w = 1.22 \frac{\lambda f}{R}.$$  \hspace{1cm} (1)

The PSF of our pinhole imaging system (i.e. the intensity distribution of the Airy pattern) can be modeled by the
Fraunhofer diffraction of the pinhole aperture. For a circular aperture with radius $R$, the PSF is given by [24]:

$$P(x, y) = \left( \frac{A}{\lambda z} \right)^2 \left[ \frac{2 J_1(kR\sqrt{x^2 + y^2}/z)}{kR\sqrt{x^2 + y^2}/z} \right]^2,$$

(2)

where $(x, y)$ is the spatial coordinate on the observation plane, $A = \pi R^2$ is the area of the aperture, $z$ is the distance from the aperture to the observation plane, $J_1$ is a Bessel function of the first order and the first kind, and $k = 2\pi/\lambda$ is the wavenumber where $\lambda$ is the light wavelength. The Fraunhofer approximation is usually only valid for far-field observations, i.e. at a distance $z \gg R^2/\lambda$ which is satisfied by our pinhole camera system.

The forward model formulates how an image is formed from an object or scene by accounting for the optical response of the camera from the PSF. Specifically, the image is the result of a convolution of the object and the PSF, represented as,

$$i(x, y) = o(x, y) \ast P(x, y),$$

(3)

where the image $i$, object $o$, and PSF are denoted as functions of position $(x, y)$ in the spatial domain. It’s also important to note that, in practice, the image sensor introduces noise that is added to the final image and can be accounted for by adding the noise model to Eq.3, i.e. $i = o \ast P + \eta$.

This forward model is extensively used in our paper including to help create synthetic data for training our network as well as a reblur loss to check the consistency of our network’s image restoration with given camera measurements.

### 3.2. Denoise Module

Due to the small aperture of a pinhole, the resulting captured images will have a high amount of noise with photon shot noise dominating [63, 31]. Thus our first step in our pipeline (after demosaicing the RAW image), is to perform denoising. One big advantage to the blur caused by the pinhole is that it creates a global frequency limit to the true image intensity, and thus any signal above the PSF’s frequency cutoff is attributable to noise. We observed in practice that even a simple, ideal low-pass filter with the same frequency cutoff used for denoising gave high-quality results when the image was passed to subsequent deblurring.

However, since noise can also occur at lower and middle frequencies, we utilize a denoise network based on the FFDNet architecture [70] to further mitigate the noise. Our choice of FFDNet was informed by an ablation study of various neural network architectures which we present in Sec. 5. As discussed in [59], FFDNet’s unique and simple initial pixel-shuffling down-sample layer doubles the receptive field without increasing the network size, resulting in faster execution time and smaller memory footprint, while being able to handle various noise types. To train this architecture, we utilize synthetic data and simulate photon and Gaussian read noise for various ISOs for the camera as detailed in Sec. 4.

### 3.3. Deblur Module

Once the denoised image is obtained from the denoise module, we utilize the state-of-the-art DeblurGANv2 [38] to restore the blurry pinhole image, given its superior performance in our ablation study in Sec. 5. Since the original model is trained primarily for motion blur, the pre-trained model is not suitable for the optical blur in pinhole images. For our pipeline, we train this network on our own synthesized pinhole datasets for it to perform deblurring on pinhole images.

### 3.4. Reblur Loss

One key challenge to a joint denoising and deblurring architecture is that deblurring can undo the effects of denoising and reintroduce noise-like artifacts into the final image. This is particularly the case in background patches in the image, where the deblur module needs to infer the missing data without many textures or patterns to infer from.

To alleviate this, we introduce joint training of our network using a reblur loss. Reblur losses have been introduced in other contexts including motion blur [14, 43] and lensless imaging [44, 48]. We utilize the captured PSF from our camera system and perform the convolution of our generated output from the network to form an estimated reblurred image $\hat{i}$. Then we utilize an MSE loss between $i$ and the original lensless image $i$ to fine-tune our network during training. Thanks to the simplicity of the pinhole imaging model, the PSF convolution reblur model is highly accurate and can be measured easily, making the reblur loss more effective than motion blur cases [43, 14].

### 4. Implementation

#### 4.1. Training Dataset

To train our network architecture, we use the HDR+ subset [28] containing, 153 images of 4048 × 3036 resolution. These are trained with 256 × 256 size patches. We create a simulated blurred version of this dataset as our input during training by convolving the real captured PSF with the original data. The original data is then used as ground truth while training the network. We implement both photon shot noise and Gaussian read noise in our simulator, using a realistic noise simulator with parameters set for various ISOs which we randomly toggle during training.

#### 4.2. Implementation details.

Our training procedure consisted of the following: (1) separately training the denoise module on blurred images
with noise, and the deblur module with the subsequent de-noised output images. Both the FFDNet and DeblurGANv2 are initially trained from scratch on the synthetic dataset. (2) After separate training, joint finetuning of the architecture is done while incorporating the reblur loss.

Our system was trained on two NVIDIA 2080Ti GPUs. While training, we use a batch size of 1 image which is cropped to multiple patches of size 256 × 256 per iteration. For both, the generator and discriminator, we use the Adam [34] optimizer using a learning rate of $1e^{-4}$ for 50 epochs then reduce the learning rate with a linear decay from $1e^{-4}$ to $1e^{-7}$ epochs for another 200 epochs. For the generator loss, we assign weights to $L_{MSE}$, $L_{perc.}$, and $L_{adv}$ of $\lambda_{MSE} = 0.5$, $\lambda_{perc.} = 0.006$, and $\lambda_{adv} = 1e^{-3}$ which is used for both the generator and discriminator adversarial losses.

4.3. Real data capture

Additionally, we capture some real pinhole camera images which we use for both quantitative and qualitative analysis of our pipeline. These are captured using a Panasonic Lumix G85 mirrorless camera with a micro 4/3 sensor and the Thingyfy Pinhole Pro using a pinhole diameter of 0.15 mm. The images are captured at 1/30 second exposure time with 3200 ISO. From the RAW image, we only apply black level correction and demosaicing to keep the tone-mapping linear.

For our real data capture, it is not trivial to capture ground truth images. Even using a beamsplitter to optically align a camera with a lens with the pinhole camera does not suffice, as the lens image will have optical aberrations and a finite depth-of-field. Thus in this paper, we choose to train on synthetic data and only use our network at inference time on real data. This has the distinct advantage of showing the generalizability of our network. For real data, due to a possible mismatch between real noise and our noise simulations, we lowpass filter the images with the PSF before running through our denoise and deblur modules. We found this helps improve results for our network as the denoising network is expecting noise statistics similar to that of the synthetic data.

To compare against other lensless camera systems, namely the FlatCam and DiffuserCam, we also capture scenes displayed from a monitor similar to these papers. We display their test dataset and capture these with our pinhole camera. Note that our network is not trained on their datasets, but only tested on their images at inference.

5. Results

5.1. End-to-end results

Tradeoff between pinhole and lenses. Smartphone photography scaling requires larger sensors, forcing lens size to scale exponentially and increase their size/weight. Lenses also have inherent aberrations (spherical, chromatic, etc.), while lensless alternatives are much more lightweight and free of those aberrations. Thus, we do not compare to lens systems, but rather, other lensless work instead. However, we did conduct a modulation transfer function (MTF) analysis shown in Figure 4. Our DCO improves the resolution in terms of MTF50 of the pinhole system by 4.2x, achieving 63.5% of the lens counterpart.

Real-world scene results. We show results on real-world scenes captured by pinhole camera and restored by our proposed pipeline in Figure 1. Note that our network has not been trained on real-world data, but solely on synthetic data simulated from the HDR dataset. As you can see in (c) and (d), the denoising and deblurring networks are respectively reducing the noise in the images and sharpening the resulting image. The resulting image in (d) is comparatively sharp as compared to a lens-camera result in (e).

We present some results from real pinhole captured low-light images in Figure 5, using the Panasonic m4/3 camera and Thingyfy Pinhole Pro. These feature a variety of scenes
with different objects and textures. Note that these images have no corresponding ground truth as we are not imaging a monitor like our lensless system comparisons. We see our captured pinhole images in (a), the resulting DCO reconstructions in (b), and comparisons to two off-the-shelf optimization methods, ADMM (c) and a simple Wiener filter (d) for deconvolution/deblurring, and (e) DMPHN [67], a modern deblurring neural network. Note how our method recovers back sharp, high-frequency details with less noise or artifacts as compared to ADMM, Wiener filter, and DMPHN. All these images were taken at 1/30 seconds exposure with ISO 3200.

Figure 5: Real captured results for pinhole camera and our DCO results. Note the spatial frequencies recovered using our network, which has not been trained on real data, but transferred from a synthetic dataset.
Figure 6: **Lensless comparisons with FlatNet.** We compare the input and reconstruction of our proposed system (DCO) with FlatCam/FlatNet [6, 33] on monitor captures of the ILSVRC 2012 dataset [50]. Note that our input and results are of higher resolution (1920 × 1080) and have higher frequency features without much reconstruction artifacts compared to the FlatNet results from [33].

**Display captured results.** Here we perform a comparison with a state-of-art, end-to-end lensless camera system, FlatNet [33]. We use the testing dataset from [33] and display the images on a monitor, and then capture it with exposure time 1/30s at ISO 3200 with our hardware pinhole camera. In Figure 6, we show the qualitative results of this comparison. Note that the pinhole naturally captures less light than FlatCam due to its single aperture. This results in a higher noise level for our DCO reconstruction in (b). However, note that the details of the image are preserved at high spatial frequencies as compared to FlatNet. In FlatNet’s reconstructions (d), certain artifacts from the machine learning include warping along the castle’s reflection, distorted signs and text, and non-physically realistic deformations.

In Figure 7, we compare the results from our proposed lensless system (DCO) with a phase mask-based lensless imaging system, DiffuserCam [44]. Similar to DiffuserCam, we display the ground-truth images on an LCD monitor, capture the images using the lensless pinhole camera, and process them using the DCO framework. Note that DiffuserCam’s raw images are captured at 1920 × 1080, reconstructed at 480 × 270, and cropped to 380 × 210. For their evaluation, the reference images are captured using a lensed camera through a beam splitter, then resized and cropped to the same 480 × 270 size. In contrast, our raw captures and reconstructions are all conducted at 1920 × 1080, the reference images in our evaluation are the original displayed images, resized to 1920 × 1080. Our results show that pinhole cameras are a viable alternative to coded aperture cameras for traditional 2D photography applications.

In Table 2, we report our PSNR and SSIM numbers as compared to FlatNet and DiffuserCam as reported by those papers [33, 44]. We score higher in PSNR than both methods by a small margin which shows the advantage of our opinion. Note that an SSIM metric evaluation was not available in DiffuserCam’s paper.

|               | FlatCam | DiffuserCam | DCO/Pinhole |
|---------------|---------|-------------|-------------|
| PSNR (dB)     | 19.62   | 21.31       | 22.31       |
| SSIM          | 0.64    | N/A         | 0.634       |
Figure 7: **Lensless comparison with DiffuserCam.** We compare the input and reconstruction of our proposed system (DCO) with DiffuserCam, a phase mask-based lensless imaging system, on the DiffuserCam datasets. Note that though our input and results are of higher resolution (1920 × 1080) compared to the DiffuserCam results from [44].

**Video Reconstruction.** Finally, we show a real-world video captured from a monitor at 1/30 seconds exposure with our pinhole camera. Note that since our current hardware cannot support video capture of RAW frames, we captured each frame individually by using a synchronized shutter. As one can see from the video frames, the input pinhole images are blurry and suffer from noise effects. Our method improves sharpness around object edges and enhanced textures in the wood, without noticeable noise degradation.

We do note some limitations to applying our network to video frames. The resulting video suffers from some temporal artifacts including flickering. While there is still room to improve upon video results, we still feel this demonstrates the improvement benefits of our DCO network in practice.

5.2. Ablation studies

**Denoise module.** Here we benchmark three denoise network candidates for denoising performance on the synthetic pinhole noise data. Overall, FFDNet performs better than FOCNet, and is on par with FC-AIDE quantitatively but with improved perceptual quality. Therefore, we chose FFDNet for our final pipeline based on this ablation study.

|            | FOCNet | FC-AIDE | FFDNet |
|------------|--------|---------|--------|
| PSNR (dB)  | 29.2964| 30.0940 | 30.0915|
| SSIM       | 0.7919 | 0.9448  | 0.9452 |

**Deblur module.** We also performed an ablation study to compare three different deblurring networks for our pipeline on the synthetic pinhole blur data. These networks were DMPHN [67], SIUN [65], and DeblurGANv2 [38]. As we can see from the table, DeblurGANv2 achieved the highest scores and thus was our choice for the deblurring module in our pipeline.

|            | DMPHN  | SIUN   | DeblurGANv2 |
|------------|--------|--------|--------------|
| PSNR (dB)  | 27.273 | 26.234 | 27.385       |
| SSIM       | 0.6805 | 0.6372 | 0.6827       |
We captured data with the ISO fixed as 1600 and perform the reconstruction with the proposed DCO framework. Top row is input, middle row is brightened input, and the last row is DCO reconstruction. The results indicate that light throughput does have a noticeable impact on the final reconstruction quality.

We captured data with the exposure time fixed as 1/30 seconds and perform the reconstruction with the proposed DCO framework. Top row is input, middle row is brightened input, and the last row is DCO reconstruction. The results indicate that the ISO setting does not impact the final reconstruction results significantly.

Exposure time. At the very low light throughput condition the pinhole camera operates in, the noise is dominantly photon noise. To understand photon noise’s impact on the reconstruction, we fixed the digital gain setting of the camera to be ISO 6400 and conduct reconstruction at various exposure times. The results are shown in Figure 9. We do observe that reconstruction, particularly denoising, gets more challenging when light throughput decreases.

ISO. In contrast to the exposure sweep experiment, we also fixed the exposure time to be 1/30 seconds and sweep across different ISO settings from 1600 to 25600. In this case, the light throughput is constant, and read noise is amplified variously depending on the ISO settings. As Figure 10 shows, we found ISO settings do not impact the final reconstruction results noticeably.

6. Discussion

In this paper, we presented a practical pinhole photography pipeline to restore low-light images. To correct the pinhole camera issues of blur and high-noise from low-light imaging, the pipeline performs denoising using a low-pass filter with a frequency limit prior based on the optical point spread function and then passing through a deep learning-based denoising network FFDNet [70] and deblurring using a retrained DeblurGANv2 [38] module on our synthesized pinhole dataset. We verify our pipeline on test data from the synthesized HDR+ dataset as well as real-world pinhole captured images to show qualitative improvements for both denoise and deblur problems. We also note that there is still room for improvements to the deblurring results, particularly for high-frequency features.
We plan to extend the work in this paper to include a modified video-specific pipeline as well as options for a learnable low-pass filter. There also may be further avenues for applications to work on past practical pinhole photography for 2D images. Since the exposure of the image can be fixed, HDR imaging for pinhole cameras can be an interesting direction that can further its potential for smartphone photography. Additionally, stereo pinhole photography can open up use cases for 3D depth estimation for SLAM algorithms and robotics in general.

References

[1] Abdelrahman Abdelhamed, Marcus A Brubaker, and Michael S Brown. Noise flow: Noise modeling with conditional normalizing flows. In Proceedings of the IEEE/CVF International Conference on Computer Vision, pages 3165–3173, 2019. 4

[2] Abdullah Abouaaim and Michael S Brown. Defocus deblurring using dual-pixel data. In European Conference on Computer Vision, pages 111–126. Springer, 2020. 4

[3] Nick Antipa, Grace Kuo, Reinhard Heckel, Ben Mildenhall, Emrah Bostan, Ren Ng, and Laura Waller. Diffusercam: lensless single-exposure 3d imaging. Optica, 5(1):1–9, 01 2018. 2, 3, 4

[4] Nick Antipa, Grace Kuo, Reinhard Heckel, Ben Mildenhall, Emrah Bostan, Ren Ng, and Laura Waller. Diffusercam: lensless single-exposure 3d imaging. Optica, 5(1):1–9, 2018. 2, 3

[5] M Salman Asif, Ali Ayremlou, Aswin Sankaranarayanan, Ashok Veeraraghavan, and Richard G Baraniuk. Flatcam: Thin, lensless cameras using coded aperture and computation. IEEE Transactions on Computational Imaging, 3(3):384–397, 2016. 3

[6] M. S. Asif, A. Ayremlou, A. Sankaranarayanan, A. Veeraraghavan, and R. G. Baraniuk. Flatcam: Thin, lensless cameras using coded aperture and computation. IEEE Transactions on Computational Imaging, 3(3):384–397, 09 2017. 2, 3, 4, 8

[7] Amir Beck and Marc Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. SIAM Journal on Imaging Sciences, 2(1):183–202, 2009. 4

[8] Vivek Boominathan, Jesse Adams, Jacob Robinson, and Ashok Veeraraghavan. Phlatcam: Designed phase-mask based thin lensless camera. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020. 3

[9] Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, and Jonathan Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. Foundations and Trends® in Machine learning, 3(1):1–122, 2011. 4

[10] Antoni Buades, Bartomeu Coll, and J-M Morel. A non-local algorithm for image denoising. In 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’05), volume 2, pages 60–65. IEEE, 2005. 4

[11] Patrizio Campisi and Karen Egiazarian. Blind image deconvolution: theory and applications. CRC press, 2017. 4

[12] Ayan Chakrabarti. A neural approach to blind motion deblurring. In European conference on computer vision, pages 221–235. Springer, 2016. 4

[13] Chen Chen, Qifeng Chen, Jia Xu, and Vladlen Koltun. Learning to see in the dark. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 3291–3300, 2018. 4

[14] Huaijin Chen, Jinxue Gu, Orazio Gallo, Ming-Yu Liu, Ashok Veeraraghavan, and Jan Kautz. Reblur2deblur: Deblurring videos via self-supervised learning. In 2018 IEEE International Conference on Computational Photography (ICCP), pages 1–9. IEEE, 2018. 4, 5

[15] Sung Hyun Cho and Seung Yong Lee. Fast motion deblurring. In ACM SIGGRAPH Asia 2009 papers, pages 1–8. 2009. 4

[16] Leonardo Da Vinci. Codex atlanticus. Biblioteca Ambrosiana, Milan, 26(1), 1894. 3

[17] Kostadin Dabov, Alessandro Foi, Vladimir Katkovnik, and Karen Egiazarian. Image denoising by sparse 3-d transform-domain collaborative filtering. IEEE Transactions on Image Processing, 16(8):2080–2095, 2007. 4

[18] Paul Debevec and Jitendra Malik. Recovering high dynamic range radiance maps from photographs. In Proceedings of the 24th annual conference on Computer graphics and interactive techniques, pages 369–378, 1997. 4

[19] Giambattista Della Porta. Magia naturalis, volume 2. 1714. 3

[20] Guillaume Druart, Nicolas Guérit, Jean Taboury, Sylvain Rommeluère, Riad Häïdar, Jérôme Primot, Manuel Fendler, and Jean-Charles Cigna. Compact infrared pinhole fisheye for wide field applications. Appl. Opt., 48(6):1104–1113, Feb 2009. 3

[21] Thomas Eboli, Jian Sun, and Jean Ponce. End-to-end interpretable learning of non-blind image deblurring. In ECCV 2020-16th European Conference on Computer Vision, 2020. 4

[22] Patrick Robert Gill, Changhyuk Lee, Albert Wang, and Alyosha Molnar. A Microscope camera using direct fourier-domain scene capture. Optics Letters, 36(15):2949–2951, 2011. 3

[23] Patrick R Gill and David G Stork. Lensless ultra-miniature imagers using odd-symmetry spiral phase gratings. In Computational Optical Sensing and Imaging, pages CW4C–3. Optical Society of America, 2013. 3

[24] Joseph W Goodman. Introduction to Fourier optics. Roberts and Company Publishers, 2005. 3, 5

[25] Shi Guo, Zifei Yan, Kai Zhang, Wangmeng Zuo, and Lei Zhang. Toward convolutional blind denoising of real photographs. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 1712–1722, 2019. 4

[26] John H. Hammond. The Camera Obscura, A Chronicle. Taylor & Francis, 1981. 3

[27] Arthur Cobb Hardy, Fred Hiram Perrin, et al. Principles of Optics, pages 124–126. McGraw-Hill Book Company, Inc., 1932. 3

[28] Sam Hasinoff, Dillon Sharlet, Ryan Geiss, Andrew Adams, Jonathan T. Barron, Florian Kainz, Jiawen Chen, and Marc
Levoy. Burst photography for high dynamic range and low-light imaging on mobile cameras. SIGGRAPH Asia, 2016. 4, 5

[29] Matthew Hirsch, Sriram Sivaramakrishnan, Suren Jayasuriya, Albert Wang, Alyoasha Molnar, Ramesh Raskar, and Gordon Wetzstein. A switchable light field camera architecture with angle sensitive pixels and dictionary-based sparse coding. In 2014 IEEE International Conference on Computational Photography (ICCP), pages 1–10. IEEE, 2014. 3

[30] O.P. Ivanov, A.N. Sudarkin, V. E. Stepanov, and L.I. Urutskoev. Portable x-ray and gamma-ray imager with coded mask: performance characteristics and methods of image reconstruction. Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, 422(1):729–734, 1999. 3

[31] Suren Jayasuriya. Image Sensors, page 1–5. Springer International Publishing, 2 edition, 2021. 5

[32] Salman S Khan, VR Adarsh, Vivek Boominathan, Jasper Tan, Ashok Veeraraghavan, and Kaushik Mitra. Towards photorealistic reconstruction of highly multiplexed lensless images. In Proceedings of the IEEE International Conference on Computer Vision, pages 7860–7869, 2019. 4

[33] Salman Siddique Khan, Varun Sundar, Vivek Boominathan, Ashok Veeraraghavan, and Kaushik Mitra. Flatnet: Towards photorealistic scene reconstruction from lensless measurements. IEEE Transactions on Pattern Analysis and Machine Intelligence, page 1–1, 2020. 8

[34] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. arXiv 1412.6980, 2014. 6

[35] Rudolf Kingslake. Lenses in Photography. A. S. Barnes and Co., 1963. 3

[36] Deepa Kundur and Dimitrios Hatzinakos. Blind image deconvolution. IEEE signal processing magazine, 13(3):43–64, 1996. 4

[37] Orest Kupyn, Volodymyr Budzan, Mykola Mykhailych, Dmytro Mishkin, and Jiří Matas. Deblurgan: Blind motion deblurring using conditional adversarial networks. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 8183–8192, 2018. 4

[38] Orest Kupyn, Tetiana Martyniuk, Junru Wu, and Zhangyang Wang. DeblurGAN-v2: Deblurring (orders-of-magnitude) faster and better. In The IEEE International Conference on Computer Vision (ICCV), Oct 2019. 4, 5, 9, 10

[39] Marc Lebrun, Antoni Buades, and Jean-Michel Morel. A nonlocal bayesian image denoising algorithm. SIAM Journal on Imaging Sciences, 6(3):1665–1688, 2013. 4

[40] Shuai Li, Mo Deng, Justin Lee, Ayan Sinha, and George Barbastathis. Imaging through glass diffusers using densely connected convolutional networks. Optica, 5(7):803–813, 2018. 4

[41] Yunzhe Li, Yujia Xue, and Lei Tian. Deep speckle correlation: a deep learning approach toward scalable imaging through scattering media. Optica, 5(10):1181–1190, 2018. 4

[42] Orly Liba, Kiran Murthy, Yun-Ta Tsai, Tim Brooks, Tianfan Xue, Nikhil Karnad, Qurui He, Jonathan T. Barron, Dillon Sharlet, Ryan Geiss, Samuel W. Hasinoff, Yael Pritch, and Marc Levoy. Handheld mobile photography in very low light. ACM Trans. Graph., 38(6), Nov. 2019. 4

[43] Tomer Michaeli and Michal Irani. Blind deblurring using internal patch recurrence. In European Conference on Computer Vision, pages 783–798. Springer, 2014. 4, 5

[44] Kristina Monakhova, Joshua Yurtsever, Grace Kuo, Nick Antipa, Kyrollos Yanny, and Laura Waller. Learned reconstructions for practical mask-based lensless imaging. Opt. Express, 27(20):28075–28090, Sep 2019. 4, 5, 8, 9

[45] Mukesha C Motwani, Mukesha C Gadiya, Rakhi C Motwani, and Frederick C Harris. Survey of image denoising techniques. In Proceedings of GSPX, volume 27, pages 27–30, 2004. 4

[46] Shree K Nayar and Moshe Ben-Ezra. Motion-based motion deblurring. IEEE Transactions on Pattern Analysis and Machine Intelligence, 26(6):689–698, 2004. 4

[47] Phillip A. Newman and Vincent E. Ribble. Pinhole array camera for integrated circuits. Appl. Opt., 5(7):1225–1228, Jul 1966. 3

[48] Joshua D Rego, Karthik Kulkarni, and Suren Jayasuriya. Robust lensless image reconstruction via psf estimation. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pages 403–412, 2021. 4, 5

[49] Leonid I Rudin, Stanley Osher, and Emad Fatemi. Nonlinear total variation based noise removal algorithms. Physica D: Nonlinear Phenomena, 60(1-4):259–268, 1992. 4

[50] Olga Russakovska, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, and Li Fei-Fei. ImageNet Large Scale Visual Recognition Challenge. International Journal of Computer Vision (IJCV), 115(3):211–252, 2015. 8

[51] Pinaki Sarder and Arye Nehorai. Deconvolution methods for 3-d fluorescence microscopy images. IEEE Signal Processing Magazine, 23(3):32–45, 2006. 4

[52] Kazuo Sayanagi. Pinhole imagery*. J. Opt. Soc. Am., 57(9):1091–1098, Sep 1967. 3

[53] Ayan Sinha, Justin Lee, Shuai Li, and George Barbastathis. Lensless computational imaging through deep learning. Optica, 4(9):1117–1125, 09 2017. 4

[54] Peter Sturm. Pinhole Camera Model, pages 610–613. Springer US, Boston, MA, 2014. 3

[55] Richard E. Swing and Dennis P. Rooney. General transfer functions of Computer Vision. Proceedings of GSPX, pages 57–64, 1966. 3

[56] Kazuyuki Tajima, Takeshi Shimano, Yusuke Nakamura, Mayu Sao, and Taku Hoshizawa. Lensless light-field imaging on mobile cameras. ACM Trans. Graph., pages 783–798, 2016. 4

[57] J. Tan, L. Niu, J. K. Adams, V. Boominathan, J. T. Robinson, R. G. Baraniuk, and A. Veeraraghavan. Face detection and verification using lensless cameras. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pages 403–412, 2021. 4, 5

[58] Jun Tanida, Tomoya Kumagai, Kenji Yamada, Shigehiro Miyatake, Kouichi Ishida, Takashi Morimoto, Noriyuki Kondou, Daisuke Miyazaki, and Yoshiki Ichioka. Thin obser-
vation module by bound optics (tombo): concept and experimental verification. *Applied Optics*, 40(11):1806–1813, 2001.

[59] Matias Tassano, Julie Delon, and Thomas Veit. An analysis and implementation of the ffdnet image denoising method. *Image Processing On Line*, 9:1–25, 2019.

[60] Antonio Torralba and William T Freeman. Accidental pinhole and pinspeck cameras: Revealing the scene outside the picture. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pages 374–381. IEEE, 2012.

[61] Antonio Torralba and William T. Freeman. Accidental pinhole and pinspeck cameras. *Int. J. Comput. Vision*, 110(2):92–112, Nov. 2014.

[62] Yicheng Wu, Vivek Boominathan, Huajjin Chen, Aswin Sankaranarayanan, and Ashok Veeraraghavan. Phasecam3d—learning phase masks for passive single view depth estimation. In *2019 IEEE International Conference on Computational Photography (ICCP)*, pages 1–12. IEEE, 2019.

[63] Li Xu, Shicheng Zheng, and Jiaya Jia. Unnatural l0 sparse representation for natural image deblurring. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1107–1114, 2013.

[64] David X. D. Yang and Abbas El Gamal. Comparative analysis of SNR for image sensors with enhanced dynamic range. In Morley M. Blouke and George M. Williams Jr., editors, *Sensors, Cameras, and Systems for Scientific/Industrial Applications*, volume 3649, pages 197 – 211. International Society for Optics and Photonics, SPIE, 1999.

[65] Minyuan Ye, Dong Lyu, and Gengsheng Chen. Scale-iterative upscaling network for image deblurring. *IEEE Access*, 8:18316–18325, 2020.

[66] M. Young. Pinhole optics. *Appl. Opt.*, 10(12):2763–2767, Dec 1971.

[67] Hongguang Zhang, Yuchao Dai, Hongdong Li, and Piotr Koniusz. Deep stacked hierarchical multi-patch network for image deblurring. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5978–5986, 2019.

[68] Kai Zhang, Wangmeng Zuo, Yunjin Chen, Deyu Meng, and Lei Zhang. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE transactions on image processing*, 26(7):3142–3155, 2017.

[69] Kai Zhang, Wangmeng Zuo, Shuhang Gu, and Lei Zhang. Learning deep cnn denoiser prior for image restoration. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3929–3938, 2017.

[70] Kai Zhang, Wangmeng Zuo, and Lei Zhang. Ffdnet: Toward a fast and flexible solution for cnn-based image denoising. *IEEE Transactions on Image Processing*, 27(9):4608–4622, 2018.

[71] Changyin Zhou, Stephen Lin, and Shree K Nayar. Coded aperture pairs for depth from defocus and defocus deblurring. *International journal of Computer Vision*, 93(1):53–72, 2011.

[72] Changyin Zhou and Shree Nayar. What are good apertures for defocus deblurring? In *2009 IEEE International Conference on Computational Photography (ICCP)*, pages 1–8. IEEE, 2009.