Designing Display Pixel Layouts for Under-Panel Cameras

Anqi Yang, *Student Member, IEEE,* and Aswin C. Sankaranarayanan

Abstract—Under-panel cameras provide an intriguing way to maximize the display area for a mobile device. An under-panel camera images a scene via the openings in the display panel; hence, a captured photograph is noisy as well as endowed with a large diffractive blur as the display acts as an aperture on the lens. Unfortunately, the pattern of openings commonly found in current LED displays are not conducive to high-quality deblurring. This paper redesigns the layout of openings in the display to engineer a blur kernel that is robustly invertible in the presence of noise. We first provide a basic analysis using Fourier optics that indicates that the nature of the blur is critically affected by the periodicity of the display openings as well as the shape of the opening at each individual display pixel. Armed with this insight, we provide a suite of modifications to the pixel layout that promote the invertibility of the blur kernels. We evaluate the proposed layouts with photomasks placed in front of a cellphone camera, thereby emulating an under-panel camera. A key takeaway is that optimizing the display layout does indeed produce significant improvements.

Index Terms—Computational Photography, Under-panel cameras, Deblurring

Fig. 1. **Improvements gained by redesigning the layout of the display pixels.** Shown above are deblurred images from three lab prototypes of under-panel cameras corresponding to (from left to right) TOLED, POLED and proposed display layouts. The insets beside each result shows the corresponding input captured photographs as well as zoomed in regions. All results emulate displays with a resolution of 300 dots per inch. The reader is encouraged to use the zoom tool to explore all three photographs.

1 Introduction

Under-panel cameras provide a way to maximize the display area on a cellphone. This provides a seamless display without the wastage associated with the bezel or potential distractions such as a "notch" and a "hole punch", thereby enhancing the aesthetics of the device.

The aesthetics achieved by placing the camera beneath the display, however, also degrades the quality of the captured photographs in two distinct ways. First, a significant portion of the incident light from a scene is blocked by the display. In many existing devices, as much as three-fourth of the light is blocked [1] and so the signal to noise ratio (SNR) of the captured photograph can be quite low, except perhaps for the brightest of scenes. Second, in addition to reducing the light levels, the display also acts as an aperture and induces a diffractive blur on the captured measurements. For the OLED displays used currently in mobile devices, this blur can have a significant spread in hundreds of pixels. Deblurring under such a large blur, especially at low SNRs, is extremely challenging.

This paper aims to redesign the display layout and, in particular, the pattern of openings through which the under-panel camera images the scene. Our goal is to shape the blur point spread function (PSF) so as to improve the conditioning of the ensuing deblurring problem. A basic result from Fourier optics suggests that the PSF observed is the squared magnitude of the scaled Fourier transform of the aperture pattern. Specializing this result to under-panel cameras, we show that the periodicity of the pixel layout as well as the specific opening at each display pixel are important factors that determine the robustness of the blur PSF to inversion.

- *The authors are with the Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA 15213.*

Armed with the insights gleaned from Fourier analysis of under-panel cameras, we introduce two key variations in the display layout. First, we argue that avoiding the periodic tiling of the display pixels, and replacing it with a random tiling, whose specifics we describe later, has the effect of reducing anisotropy of the blur. Second, optimizing the shape of the display opening found at a single display pixel can further improve the invertibility of the blur PSF; this optimized pattern is randomly tiled to create the display layout. Together, these innovations provide a rich design space for engineering PSFs that are superior to popularly used P-OLED and T-OLED displays.

Contributions. This paper advances under-panel camera technology by providing optimized display patterns that improve the quality of restored photographs. In this regard, we make the following contributions.

- *Analysis of the PSF.* Using Fourier optics, we analyze the properties of PSF of a camera under a typical OLED display, and connect its spread, periodicity and falloff to the repetitiveness as well as shape of the display openings.
- *Improving PSF conditioning via random tiling.* We propose a simple modification to display layout in the form of a random tiling where each of its pixels is randomly flipped or rotated by 90°. We provide detailed theoretical analysis of this random tiling and show that it improves the robustness of the PSF to inversion.
- *Improving PSF conditioning via optimization.* Finally, we improve the conditioning of the PSF by optimizing the shape of the opening at a pixel, which is kept the same across the display except for the random tiling. We explore two distinct approaches to achieving this: first, optimizing the invertibility of the PSF, and second, end-to-end optimization based on reconstruction performance over a collection of images.

Our proposed layouts are optimized with pre-determined constraints on the display LEDs, in terms of their size and pitch and hence, in principle, are realizable with appropriate redesign of the power/control circuitry. The contributions above are analyzed in simulations as well as real results captured from a lab prototype. The code and dataset for this work is publicly available [2]. Figure 1 shows an example of the improvements that are achieved with this redesign of the display layout. We can immediately observe that the quality of the deblurred photograph is significantly enhanced with the optimized display layout.

Limitations. The contributions above come with certain limitations. Perturbing display layout, especially breaking the periodicity, is likely to complicate the fabrication of the display and the design of the power and control wiring; but given the maturity of CMOS fabrication, we expect this to be an engineering challenge and not a fundamental limitation.

2 Related Work

Computational displays. There is a rich literature on designs that seek to enhance the capabilities of a display [3]; we focus on those that integrate a display with a camera. Early work in this space focuses on camera-projector systems that capture images from behind semi-transparent projection screens. TouchLight [4] uses stereo cameras to track gestures and DepthTouch [5] places a depth camera behind a projection screen to sense objects at different depths. BiDiScreen [6] instead interlaces image diodes into a thin liquid-crystal display and utilizes the display as a pinhole array to form a lensless imager. However, their design requires displays that have low pixel densities and, further, the imaging quality is significantly worse than webcams and cellphone cameras.

Design of LED displays and under-panel cameras. OLED displays do not require a backlight panel and can transmit light through their transparent substrates and cathodes [1]; this raises the potential for acquiring high-resolution photographs by placing cameras "under" the display panel without sacrificing the quality of displays [7], [8], [9]. In particular, there has been a focus on using Transparent-OLED (T-OLED) and Pentile-OLED (P-OLED), displays that are commonly used in commercial televisions and cellphones. Cheng et al. [7] place a camera behind a T-OLED screen, simulating the point spread function (PSF), and demonstrate that the image quality is significantly worsened due to the diffractive blur introduced by the screen.

Image restoration for under-panel cameras. Recent interest in deploying under-panel cameras in smartphones has spurred interest in techniques that can restore images captured with them. In particular, many recent techniques [8], [10], [11] use deep neural networks to handle the large blur and low SNR in under-panel imagery. To recover high-quality images, Zhou et al. [8] exploit convolutional neural networks to deblur and denoise the images captured under both T-OLED and P-OLED; they demonstrate that a deep neural network, with a UNet architecture [12] model, produces deblurred photographs that are significantly better than simple techniques like Wiener deconvolution. Emerton et al. [9] propose to tackle degradations from diffraction using structured light with specialized frequencies to illuminate the target scenes; the need for control of scene illumination, unfortunately, places significant constraints on the use of the device. Sundar et al. [13] deblur on low-resolutional images and uses a guided filter network to restore high-resolutioinal images. Puthussery et al. [14] use an encoder-decoder network and adds to each block multiple densely connected convolutional layers with different dilations. Similarly, the best-performing methods in [15] use a UNet architecture with dense residual blocks [16] added to each encoding and decoding unit. We refer to the network architecture as UNet-RDB, and train a model based on it for deblurring our captured photographs.

Coded apertures. The idea of using an amplitude mask to code the aperture of a lens has a long history in computational photography, including early work in lensless X-Ray and Gamma ray imaging [17]. Recent work in such *coded aperture* cameras has focused on robust estimation of depth from a single [18], [19] as well as multiple [20] images as well as estimating light fields [21] from a single coded image. Conceptually, the ideas in our paper fall firmly under this category of coded aperture cameras. The key difference, however, is in the smallest feature size in the coded aperture. Most prior works operate with openings where the smallest feature is significantly larger, often in hundreds of microns.

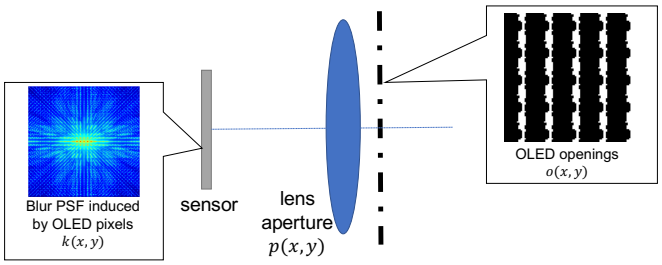


Fig. 2. Layout of the under-panel camera. The overall aperture consists of a collocated OLED display panel and a finite lens aperture.

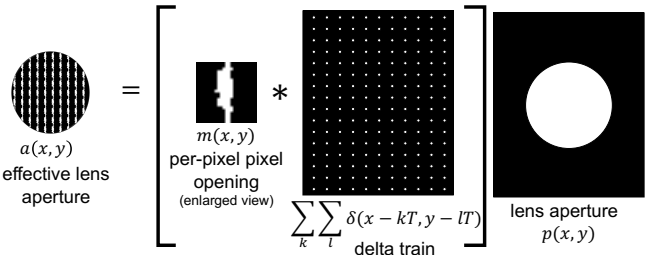


Fig. 3. Modeling the effective aperture of an under-panel camera

This permits the use of geometric optics for modeling the effect of the aperture, and further, also implies that a scene that is in focus appears sharp with little or no blur. In contrast, displays have a pixel pitch that is often smaller than 100 μm and hence, the openings have features in the scale of microns, requiring the use of tools from wave optics for modeling and analysis. This also results in a large diffractive blur even for the in-focus scene.

Coded apertures that use phase masks have also found extensive use for similar problems including extended depth of field imaging [22] as well as depth from defocus [23], [24]. The use of phase modulation requires that these techniques operate under a wave model, and, in this sense, are similar to the techniques used in this paper. However, the aperture pattern in an under-panel camera has to accommodate the OLED array, which necessarily blocks light and often in a periodic pattern; hence, we cannot model the resulting aperture as a pure phase mask.

Hence, both the design of coded apertures as well as the underlying modeling associated with prior work does not easily translate to under-panel cameras.

3 Under-panel Image Formation

In this section, we present the basics of image formation for an under-panel camera, focusing specifically on the blur PSF and its relationship to the display layout.

3.1 Derivation of the Blur PSF

Setup. Figure 2 provides the basic setup of our display-camera system. We assume that the camera lens can be well approximated as a thin lens with focal length f_0 and with an aperture given by $p(x, y)$. The display openings are described using a binary-valued function $o(x, y)$, which is assumed to be collocated with the aperture of the thin lens without any separating distance between them; this assumption greatly simplifies the derivation and is important for analytical reasoning. Finally, the incident light is assumed to be is spatially and temporally incoherent.

Spatially-invariant blur model. Let's suppose that the camera is focused on a scene at infinity. The image formed on the sensor can be written as:

$$i_{\text{blur}}(x, y) = \int_{\lambda} [i_{\text{sharp}}(x, y; \lambda) * k(x, y; \lambda)] s(\lambda) d\lambda,$$

where $*$ denotes the convolution operator, λ is the wavelength of light, $s(\lambda)$ is the camera spectral response, and $i_{\text{sharp}}(x, y; \lambda)$ is the sharp image at wavelength λ that would be formed on the sensor with an ideal thin lens, and without the display. The term $k(x, y; \lambda)$ is the blur kernel at wavelength λ, whose expressions we derive next. It is worth pointing out that the shift invariance as well as lack of interference between scene points is a consequence of the thin lens and the incoherence of light, respectively. The interested reader is referred to Chapter 3 of Goodman [25].

From basic Fourier optics, the blur kernel $k(x, y; \lambda)$ can be written as the *squared magnitude* of the scaled Fourier transform of the effective aperture function. Specifically, the effective aperture function $a(x, y)$ is the product of the lens aperture $p(x, y)$ and the display openings $o(x, y)$, i.e.,

$$a(x, y) = p(x, y)o(x, y), \tag{1}$$

then the blur PSF is given as

$$k\,(x, y; \lambda) = \left| \frac{1}{j\lambda f_0} A\left(\frac{x}{\lambda f_0}, \frac{y}{\lambda f_0} \right) \right|^2, \tag{2}$$

where $A(u, v)$ is the Fourier transformation of the $a(x, y)$.

Specializing to under-panel cameras. We now specialize the expression for the blur PSF to features commonly found in an under-panel camera. In an under-panel camera, we expect the display openings to be periodic since each display pixel is identical. Let T μm be the pixel pitch of the display; this pitch also determines the resolution of the display given as $25400/T$ dots per inch (DPI). If we denote $m(x, y)$ to be the opening pattern as pertaining to a single pixel, the overall display openings $o(x, y)$ can be constructed with copies of $m(x, y)$ repeating at a periodicity of T along both axes. As noted in Figure 3, we can mathematically express this as follows:

$$o(x, y) = m(x, y) * \sum_r \sum_c \delta(x - rT)\delta\,(y - cT)\,, \tag{3}$$

where $\delta(x)$ is the Dirac delta function. That is, the display panel is represented as the per-pixel opening $m(x, y)$ convolved with a delta train of periodicity T along both axes.

Noting that the Fourier transform of a delta train with periodicity T μm is also a delta train, but with periodicity $1/T$ $(\mu\text{m})^{-1}$ and, further, multiplication in space domain leads to convolution in Fourier domain, we can write the $A(u, v)$, the Fourier transform of the effective aperture, as

$$\begin{aligned} A(u, v) &= P(u, v) * \left[M(u, v) \sum_k \sum_l \delta\left(u - \frac{k}{T}\right) \delta\left(v - \frac{l}{T}\right) \right] \\ &= \sum_k \sum_l M\left(\frac{k}{T}, \frac{l}{T}\right) P\left(u - \frac{k}{T}, v - \frac{l}{T}\right) \end{aligned} \tag{4}$$

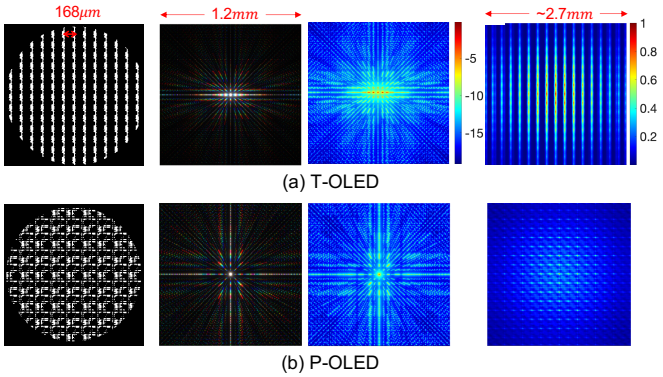


Fig. 4. Two commonly used OLED patterns, (a) T-OLED and (b) P-OLED, and the blur induced by them. For each LED type, we show (left) the display opening pattern, (center-left) the three-color tonemapped PSF, as well as (center-right) the PSF, in log-scale, corresponding to the green channel. The PSF for each color was computed by averaging across multiple wavelengths and weighted by camera spectral response. (right) The Fourier transform of the blur PSF, which is also the scaled auto-correlation of the aperture pattern.

The expression above captures the dependence of the blur PSF on all the key terms that define the under-panel camera; we analyze this dependence next.

3.2 Properties of the Blur PSF

We can analyze the blur PSF derived in (2) and (4) and derive some of its critical properties. This allows us to make the following observations about the blur PSF.

Periodic sub-structures. The blur PSF is made of repeated copies of $P(u,v)$ — scaled locally by $M(u,v)$. Once we account for the scaling by $1/(\lambda f_0)$ of $A(u,v)$, the periodicity of P is $\lambda f_0/T$. For a display with 150 DPI, lens focal length $f_0 = 10$mm and wavelength $\lambda = 0.53\mu$m, this periodicty comes to 32μm, i.e., 10-15 pixels wide on the sensor. In contrast, $P(u,v)$ once scaled by $1/\lambda f_0$ is the airy disk of the open aperture and is constrained to a few pixels. Hence, we can expect the blur PSF to have *sparse* repeating structures, each shaped like an airy disk, as seen in Figure 4.

PSF envelope and directionality. While the local structure of the PSF is shaped by P, the overall envelope of the blur kernel, that determines it spread, is primarily determined by $M(u,v)$. This implies that the per-pixel display opening $m(x,y)$ has a dominant role in shaping the blur PSF that we observe. This happens in two distinct ways. First, since $m(x,y)$ is spatially compact and restricted to be within a square of width T μm, where T is the pixel pitch, we can expect its Fourier transform $M(x,y)$ to have a spread that is inversely proportional to T. Second, directionality or anisotropy in the shape of $m(x,y)$ leads to directionality in the shape of $M(u,v)$ and, consequently, in the PSF that we observe; an example of such anisotropy can be seen in the T-OLED display in Figure 4. Directional PSF implies that we would preserve detail preferentially in some directions as opposed to others. Hence, all things considered, displays with larger pixel pitch produce smaller blur and pixel openings that are symmetric produce isotropic blur kernels.

Invertibility. We can analyze the invertibility of the blur PSF by looking at its magnitude spectra, i.e., the magnitude of its Fourier transform. Nulls and small values in the magnitude spectra are undesirable as they lead to noise amplification when we deblur the image.

Connection to the auto-correlation of the aperture. From (2), we can express the Fourier transform of $k(x,y;\lambda)$ as

$$K(u,v;\lambda) = \mathcal{AC}_a(\lambda f_0 u, \lambda f_0 v) \tag{5}$$

where $\mathcal{AC}_a$ is the auto-correlation function of $a(x,y)$; this expression comes from the fact that the power spectral density and auto-correlation are Fourier pairs, and hence $|A(u,v)|^2$ is the Fourier transform of $\mathcal{AC}_a(x,y)$. Since each color channel is a weighted sum over the visible waveband, the magnitude spectra of the blur in each color channel will be the corresponding weighted sum of $K(u,v;\lambda)$, or equivalently, $\mathcal{AC}_a(\lambda f_0 u, \lambda f_0 v)$. This smoothens the PSF as well as its Fourier transform; however, it does not change the conclusions that we draw below which are based on the monochromatic blur kernel.

Auto-correlation for periodic tiling. The auto-correlation of $a(x,y)$ depends on the lens aperture $p(x,y)$ as well as the per-pixel display opening $m(x,y)$ as given by (1) and (3). While a general expression for $\mathcal{AC}_a(x,y)$ is hard to derive, we can derive meaningful insights simply by looking at its values for small values of (x,y). Specifically, when the pitch of the display T is significantly smaller than the lens aperture, there are multiple display pixels within the aperture. In this scenario, the auto-correlation $\mathcal{AC}_a$ at small displacements (x,y) becomes repeating copies of $\mathcal{AC}_m$, the auto-correlation of $m(x,y)$, scaled by the number of copies of $m(x,y)$ within the lens aperture. Here, we directly see the effect of the per-pixel pattern $m(x,y)$ and its periodic tiling in the invertibility of the blur PSF. If $m(x,y)$ is compact along any direction, then we can expect the repeated copies of its auto-correlation to not overlap which results in nulls. Further, even if nulls are avoided, decaying tails in the $\mathcal{AC}_m$ would lead to a blur kernel that is not robust to noise. The auto-correlations associated with T/P-OLED displays are shown in Figure 4; we can clearly observe the periodic structures with peaks and nulls, as a consequence of the periodicity of the display tiling.

4 Rethinking Display Pixel Layout

We now propose new display layouts that are motivated by the analysis laid out in Section 3.2. In particular, we are interested in enabling robustly-invertible PSFs by shaping the pixel layout over the aperture.

Approach. A straightforward approach is to optimize the entire pixel layout over the lens aperture under an appropriate cost on the PSF. However, any solution has to accommodate an LED array with the appropriate resolution and fill factor/LED footprint. While it is possible, in principle, to write out this as a constrained optimization problem, we adopt a different technique that makes the display design significantly simpler.

Our proposed approach relies on two key observations.

- *Random tiling.* First, the periodic tiling of $m(x,y)$ in the display layout causes its auto-correlation to have small

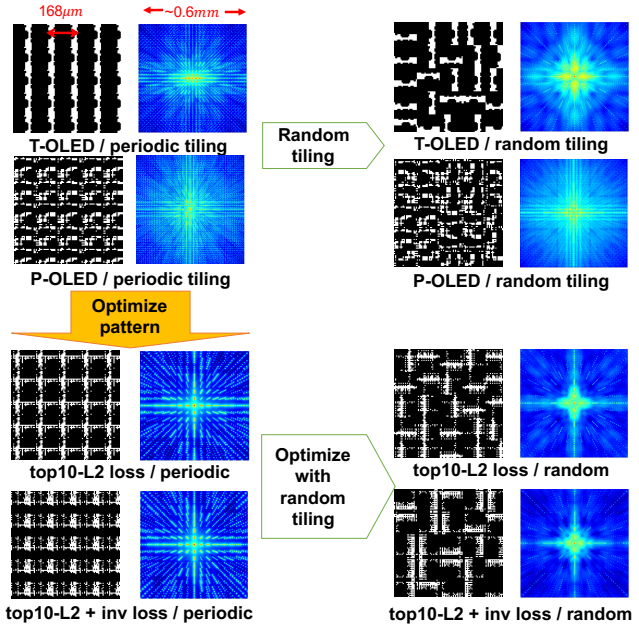


Fig. 5. We propose to optimize pixel layout by random tiling pixels and optimizing individual pixel openings. The figure above shows how the blur PSF changes when we introduce random tiling without changing the per-pixel pattern (top row), and when we optimize for per-pixel patterns under different criteria (bottom row), both with and without random tiling.

values and nulls, which results in a non-invertible blur. This can be alleviated if we tiled the display randomly, where each pixel is randomly chosen between one of many patterns.

- *Optimizing for the per-pixel pattern* $m(x, y)$. Second, we can optimize the shape of the opening at a *single* display pixel with the goal of producing a robust PSF. This pattern is subsequently tiled, with random rotations and flips, to create a random tiling over the aperture of the lens. As a result, the number of parameters to optimize is significantly smaller than what we would have if we optimized for the entire display.

This design methodology has the added advantage of relying on a single per-pixel pattern $m(x, y)$. As long as this pattern permits the LED of a certain footprint, its tiling at the desired DPI, under the random rotation and flip, ensures a feasible LED array over the aperture. Figure 5 illustrates how the PSF changes with each of the two modifications.

4.1 Random Tiling of the Display Pixel

To understand how random tiling affects the PSF, we will perform the derivation with a 1D display and sensor; the extension to 2D is straightforward and provided in the supplemental material.

1D analysis. Let's first consider a simple 1D display where each pixel is randomly chosen between one of two patterns. Let $m_1(x)$ and $m_2(x)$ be the two potential candidates at each pixel. Also suppose that there are R display pixels over the lens aperture. With this, the overall aperture function $a_r(x)$, including the lens aperture, is given as

$$a_r(x) = \sum_{k=0}^{R-1} \frac{1+U_k}{2} m_1(x-kT) + \frac{1-U_k}{2} m_2(x-kT), \quad (6)$$

where $\{U_k, 0 \leq k \leq R-1\}$ are iid Bernoulli random variables taking values in $\{+1, -1\}$ with equal probability; hence, U_k selects between m_1 and m_2 at the k-th pixel. By rearranging the terms involving U_k in (6), we get

$$a_r(x) = b_1(x) * \sum_{k=0}^{R-1} \delta(x-kT) + b_2(x) * \sum_{k=0}^{R-1} U_k \delta(x-kT), \quad (7)$$

where

$$b_1(x) = [m_1(x)+m_2(x)]/2, \text{ and } b_2(x) = [m_1(x)-m_2(x)]/2.$$

The term involving $b_1(x)$ is similar to the display model from before, namely, tiled copies of a pattern over the lens aperture. However, this is now modified by the second term whose effect can be studied next.

From basic signal processing, we can write the Fourier transform of the aperture a_r as

$$A_r(u) = B_1(u) \sum_{k=0}^{R-1} e^{-j2\pi ukT} + B_2(u) \sum_{k=0}^{R-1} U_k e^{-j2\pi ukT}.$$

When the pixel pitch T is much smaller than the lens aperture, or equivalently when R is large, the observed PSF is well approximated by the *expected* value of $|A_r(u)|^2$, which can be expressed as

$$\mathrm{E}[|A_r(u)|^2] = |B_1(u)\Delta(u)|^2 + R|B_2(u)|^2. \quad (8)$$

In essence, when we randomly tile two patterns m_1 and m_2, the expected blur PSF is the sum of two terms: the first term $|B_1(u)\Delta(u)|^2$ that corresponds to a periodic tiling of $m_1 + m_2$ over the aperture of the lens, and the second term that is simply R times the Fourier transform of $m_1 - m_2$.

Further, as before, we can analyze the Fourier transform of the blur for invertibility and robustness. Here, there are two terms: first, the auto-correlation of $(m_1 + m_2)/2$ with its periodic tiling over the aperture — this has a behavior similar to what we get with a conventional tiled display, however, $(m_1 + m_2)/2$ is more isotropic than a single m_1 or m_2; and second, the auto-correlation of $R(m_1 - m_2)/2$ — without any tiling — which stabilizes the PSF.

Extending analysis to the 2D case. To extend the analysis to the 2D display case, we need to first identify the number of patterns that we choose from. While this is something we can choose freely, there are advantages to having a single pixel layout and simply rotating / flipping it. As a consequence, there are four distinct patterns that can appear in any pixel: the unperturbed pattern, the pattern under a 90° rotation, the pattern under a flip, and finally, the pattern under both operations.

As with the 1D case, the key observation is that the PSF for a randomly tiled display is made of two terms: a term that corresponds to periodic tiling and a second term that is non-repetitive. We provide the details in the supplementary.

Evaluating the efficacy of random tiling. Figure 5 shows how the PSF of the T-OLED and P-OLED pattern changes when we subject it to random tiling. We can observe that both the periodic sub-structures as well as the anisotropy of the original PSF are reduced significantly. While we provide a detailed quantitative evaluation of random tiling in Section 5, Figures 6 and 7 show how the auto-correlation and

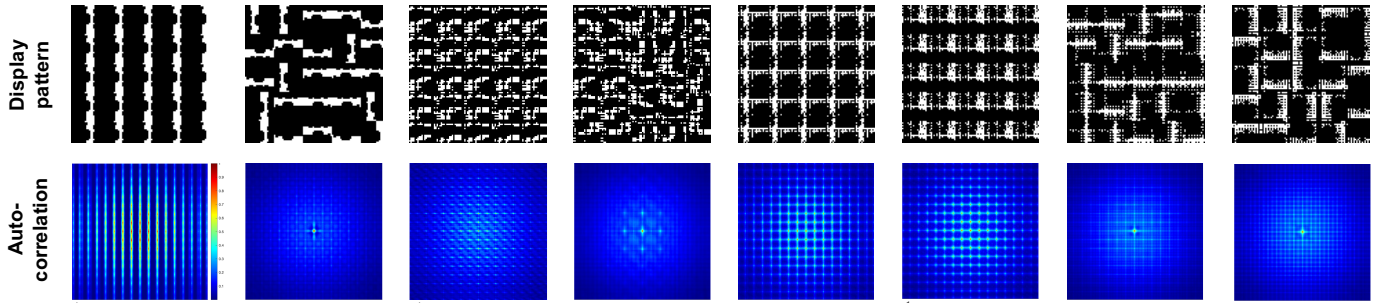
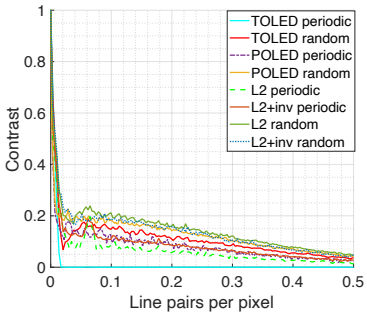


Fig. 6. **Effect of random tiling and pixel shape optimization.** In the left four columns, we show the effect of random tiling to two common displays. In the right four columns, we show our optimized pixel opening shapes from two losses, top-10 L2 and top-10L2+invertible loss, and each with two tiling strategies during optimization.



	AUC	LTR(%)
TOLED / periodic	2.43	20.7
TOLED / random	16.47	20.8
POLED / periodic	12.44	23.7
POLED / random	18.98	23.8
L2 / periodic	10.26	23.0
L2 + inv / periodic	13.51	22.7
L2 / random	21.98	22.9
L2 + inv / random	20.55	22.6

Fig. 7. **Comparison of MTF plots.** We compare radially min MTFs of different patterns and the table summarizes the area under the MTF curve (AUC) and light transmittance rate (LTR) for each mask. Larger AUC is better.

modulation transfer function (MTF) change with different display layouts; recall that 2D MTF corresponds to the amplitude of auto-correlation function. Each curve in Figure 7 is generated by taking the radially minimum values of the 2D MTF. The plots are generated for a single wavelength $\lambda = 610$nm. As expected, repeatedly tiled T-OLED has multiple null values in many frequencies and randomly tiled T-OLED lifts the nulls to small values and thus stabilizes the inversion. For P-OLED pattern, a randomly-tiled aperture induces larger minimum values for all frequencies than that of the repeatedly tiled aperture and thus is more robust to inversion. Figure 7 also provides quantitative comparison for different display layouts by summarizing the area under the radially minimum MTF curves (AUC) of each display and listing the corresponding light transmittance rate (LTR). We can clearly see that randomly-tiled displays have higher AUCs than their periodically tiled counterparts, again indicating that random tilings are more robust to inversion. We also provide an analysis for horizontal, vertical, and diagonal MTFs in the supplementary material.

4.2 Optimizing for the Per-Pixel Pattern

While random tiling provides improvement over a periodic one, we can further improve the efficacy of the blur PSF by designing the per-pixel opening $m(x, y)$. We formulate this as an optimization problem where seek a desirable PSF, as characterized by a loss/cost function, and optimize for the per-pixel pattern $m(x, y)$ that minimizes this cost. It is worth reemphasizing that we only optimize for a single pattern $m(x, y)$; the display layout is constructed using periodic or random tiling depending on the specifics of the design.

Optimization setup. We discretize the variable $m(x, y)$ into a 2D matrix $\vec{m} \in R^{N\times N}$. For a display pixel, each point takes value in $\{0, 1\}$, where 0 indicates closed regions that contains LEDs, control circuits and etc. and 1 represents pixel openings that allow light to comes through. Since optimizing over a binary-valued variable is not easily amenable to standard descent-based optimization, we relax $\vec{m}$ to be real-valued, and instead map its elements to $[0, 1]$ using a sigmoid function $g(\vec{m})$ to get a mask. The PSF $A(\cdot)$ is a function of the mapped pixel opening $g(\vec{m})$, as well as a number of fixed parameters that include the display pixel pitch T, the specifics of the random tiling (if used), and other parameters such as f_0 and λ.

We seek an optimal pixel opening $\vec{m}$ whose PSF is invertible with respect to Wiener deconvolution. We choose Wiener deblurring instead of a deep neural network (DNN) during optimization as in previous PSF-engineering works [26], [27], [28] for the following reasons. First, a simpler algorithm like Wiener deconvolution puts the emphasis entirely on the system conditioning, in terms of a mask that produces an invertible blur, so that the inadequecies of the mask are not suppressed by a powerful inverse algorithm. Second, by doing this, we also avoid being biased to specifics of the DNN that is used, and the data used in the process of training both the DNN and our technique. Finally, Wiener deconvolution is *blindingly* fast which is very helpful in the context of optimization.

Loss #1 —- PSF-induced loss. When using Wiener deconvolution, the estimated deblurred image $\widehat{I}_{\text{sharp}}$ can be written in terms of the *ground truth* image I_{sharp} as follows,

$$\widehat{I}_{\text{sharp}} = H(\vec{m}) I_{\text{sharp}},$$

where

$$H(\vec{m}) = \frac{A(g(\vec{m}))A(g(\vec{m}))^*}{|A(g(\vec{m}))|^2 + \epsilon}. \tag{9}$$

Here, $H(\vec{m})$ is the overall system frequency response that characterizes blurring and recovering the desired image and it is a function of $A(g(\vec{m}))$, the Fourier transform of PSF induced by the pixel opening matrix $g(\vec{m})$. When $H(\vec{m})$ is the identity operator, we obtain $\widehat{I}_{\text{sharp}} = I_{\text{sharp}}$ and, in theory, we can perfectly recover the sharp image. Hence, a good metric for optimization is to maximize the smallest

value of $H(\vec{m})$. In practice, instead of taking the smallest value of $H(\vec{m})$, we take the average of the smallest thirty percent of the elements of $H(\vec{m})$ to improve the robustness of optimization. We vectorize and sort the values in $H(\vec{m})$ so that $\{H(\vec{m})\}_1 \leq \{H(\vec{m})\}_2 \leq ... \leq \{H(\vec{m})\}_{N^2}$. The invertible loss is defined as

$$\mathcal{L}_{\text{inv}} = -\frac{1}{N'}\sum_{i=1}^{N'}\{H(\vec{m})\}_i.$$

We use $N' = \lceil 0.3N^2 \rceil$ in all our optimizations.

The loss function $\mathcal{L}_{\text{inv}}$ is closely related to the work of Mitra et al [29], where the performance of various computational imaging systems are analyzed; here, given a forward operator A, the term $\text{Tr}(\text{inv}(A^\top A))$ is to model and analyze system conditioning. For the convolutional model, $(A^\top A)^{-1}$ is closely related to the Wiener filter. However, there are some differences in how we define the loss function. While trace(inv$(A^\top A)$ minimizes MSE, we observe that the inverse filter is *only* unstable at a small number of Fourier coefficients, and so we only optimize the worst ten percent of filter coefficients which prioritizes worst-case performance as opposed to average.

Loss #2 — Data-driven loss. While PSF-induced loss provides a data agnostic metric, deblurring performance on actual images is often the gold metric. Hence, over a small dataset of images [15], we minimize the error between the ground-truth images and corresponding deblurred images, obtained from the Wiener deconvolution technique. Since flat regions in the image are easy to recover, the loss is dominated by the easy samples. We use hard-sample mining to penalize the largest reconstruction errors [30]. Specifically, we compute the residuals $\Delta I = I - \hat{I}$, vectorize and rank the absolute values of the residuals $|\Delta\vec{i}_r|$ in descending order, take the top 10% residuals to compute L2 loss. Top-10 L2 loss is formulated as

$$\mathcal{L}_{\text{data}} = \sum_{r=1}^{R}|\Delta\vec{i}_r|^2,$$

where R is the number of 10% elements in the current batch.

Choice of tiling. In addition to the loss functions, we also have different choices in how we tile the per-pixel pattern $m(x, y)$, that we optimize for, to create the lens aperture function. The standard tiling creates a periodic pattern by repeating the pattern $m(x, y)$ till it covers the aperture of the lens. We also have the choice of random tiling where the pattern is randomly flipped and rotated to create the aperture pattern. An important point is that the sequence of random flips and rotations are randomly generated once and fixed; at optimization time, the pattern $m(x, y)$ is optimized under this specific tiling.

Target function. Combining abovementioned losses, our target function is formulated as

$$\arg\min_{\vec{m}} \alpha_{\text{inv}}\mathcal{L}_{\text{inv}} + \alpha_{\text{data}}\mathcal{L}_{\text{data}} + \alpha_{\text{area}}\left|\vec{1}^T g(\vec{m})\vec{1}/N^2 - c\right|^2.$$

The last term constraints the total opening area of the mapped pixel opening $g(\vec{m})$ to be around target ratio c. We optimize for individual pixel $m(x, y)$ of $T = 168\mu$m in x,y directions and under the constraint that 20% of pixel region is open, i.e. $\frac{1}{T^2}\int_{x=0}^{T}\int_{y=0}^{T} m(x, y) = 0.2$, which are parameters of a typical T-OLED pixel. We discretize the pixel into 21×21 2D matrix, i.e. $N = 21$, where each element represents a dot of width and length of 8μm. We fix the focal length as $f = 10$mm, aperture as $f/2.5$ and use wavelength of 610nm, 530nm, and 470nm. We use stochastic gradient descent with learning rate 1 and optimize for 150 epochs. The pixel opening matrix m is initialized as an all-one matrix, i.e. the pixel is all open. In the first iteration, we set the area constraint as $c = 1$ and gradually decrease it by 0.05 every five epochs until $c = 0.2$.

Optimized layouts. For purposes of evaluation, we generate four distinct display layouts by thresholding the display pattern to binary values $\{0, 1\}$ and keep the target opening area, which are shown along with their PSFs in Figure 5. The four sets corresponds to two distinct loss functions — top-10 L2 loss, and top-10 L2 + invertible loss — and two kinds of tiling — periodic and random. We will visualize the corresponding optimized patterns and demonstrate their performance in the simulated and real experiment sections.

5 Simulated Experiments

To evaluate the performance of our techniques, we quantitively compare the recovered images generated by simulating capture behind different display patterns.

Simulation setup. We utilize thirty images provided in [8] validation set to generate degraded and ground-truth image pairs. The degraded image is generated by convolving a ground-truth image with our simulated PSFs and then adding shot noise and read-out noise to the blurry images according to the parameters of a typical cellphone camera. Specifically, we use a full well capacity of 15506 electrons and a standard deviation of read-out noise of 4.87 electrons. We simulate five different light levels with SNR varying from 24dB to 40dB and corresponding maximum number of electrons varying from 270 to 10000 (not exceeding full well capacity). To recover sharp images, we first denoise the degraded images with BM3D [31] and then deblur them using Wiener deconvolution. We measure the quality of deblurred images by comparing them with corresponding ground-truth images and compute PSNR and SSIM [32].

Effect of random tiling. We first look at the effect of introducing random tiling to existing display patterns without altering the shape of individual pixel openings. Figure 8 reports PSNR and SSIM numbers as a function of measurement noise levels. For the T-OLED display as well as optimized ones, introducing random tiling provides improvements in both metrics; for T-OLED this improvement is very significant due to inherent anistropy of the pattern.

Effect of pixel shape optimization. We compare existing display pixels with our optimized ones. In the last four columns in Figure 6, we show optimized patterns from two losses, top-10 L2 loss and top-10 L2 + invertible loss, and for each of them we show two tiling strategies — periodic repeating and random tiling that is chosen and fixed prior to optimization. During testing, we use corresponding tiling strategies to form display panels. During optimization, we

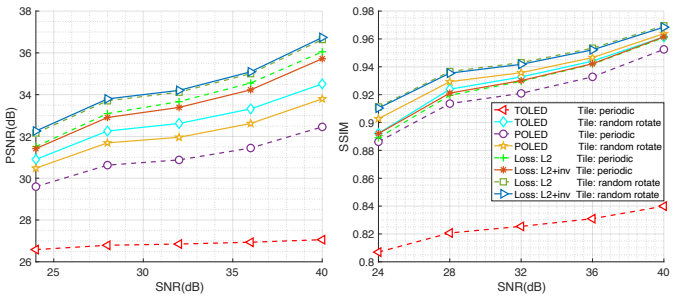


Fig. 8. **Performance of random tiling and pixel shape optimization.** We compare six display layouts on the simulated dataset and evaluate PSNR and SSIM under varying noise levels.

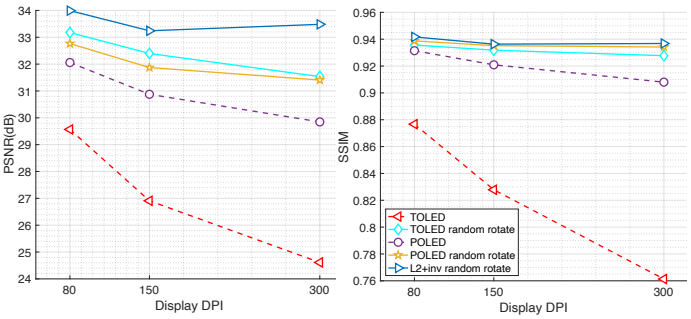


Fig. 9. **Effect of display pixel density.** We compare four pixel openings/layouts under varying display densities. The horizontal axes are display Dot-Per-Inch(DPI), and the vertical axes are PSNRs and SSIMs of the reconstructed images under a fixed noise level.

use 240 images from the training set of [8] to compute the top-10 L2 loss. As shown in Figure 8, optimizing pixel shape with periodic repeating improves the reconstruction quality by a large margin compared to the conventional T-OLED. Incorporating random tiling as described above leads to additional improvements. The pattern optimized with "top-10 L2 + invertible" loss with random tiling has the best performance and achieves more than 8dB increase in PSNR and around 0.11 increase in SSIM over T-OLED.

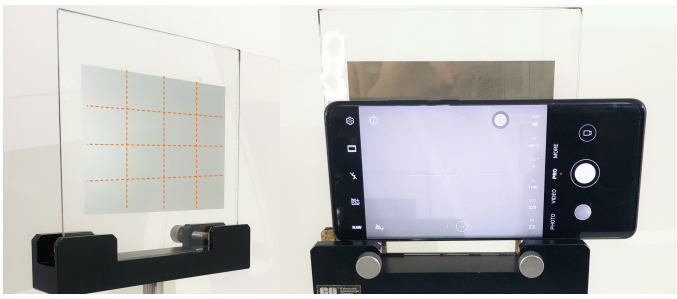
Effect of display pixel density. We show the quality of reconstructed images of conventional patterns and optimized patterns under different pixel densities. For the same pixel pattern, a higher pixel density results in a larger blur kernel and thus is harder to recover a sharp image. We fix the noise level to be SNR= 32 dB and vary pixel density from 80 DPI to 300 DPI. As seen in Figure 9, the recovered image quality decreases as pixel density increases and the optimized patterns outperform two common pixels T-OLED and P-OLED under all pixel densities. Note that the improvement on P-OLED is not as significant as on T-OLED.

Additional results. Due to space constraints, we provide additional simulation results in the supplemental material.

6 Real Experiments

We build a lab prototype to qualitatively evaluate images captured under different display designs and compare corresponding deblurring results.

Prototype. As shown in Figure 10, our prototype consists of a photolithography mask that emulates the display screen



(a) Photolithography mask (b) Under-panel camera prototype

Fig. 10. Under-panel camera lab prototype. (a) shows twelve photolighography masks that emulate different display designs. (b) shows our overall prototype where we place a cell-phone camera tightly against the printed mask and capture images by accessing the touch screen.

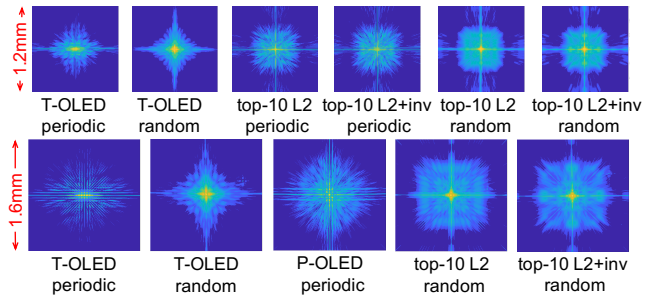


Fig. 11. We capture PSFs of different display layouts and visualize green channel in log scale. The two rows show PSFs of displays with resolutions of 150 and 300 DPI, and of size 300×300 pixels and 400×400 pixels, respectively.

and an on-the-market cellphone camera. The cellphone camera has a focal length $f = 5.56$mm and an aperture of $f/1.6$.

Camera pipeline. We extract RAW images from the camera and process them using a simple pipeline. We first radiometrically calibrate the cellphone camera and use a color checkerboard, under different lighting conditions, for white balancing and color correction. This calibration is done prior to placing the photomask in front of the lens. For each RAW image, we first demosaic, spatially downsample it to 1364×1820, denoise and deblur it, and then correct color and tonemap it to obtain the final result. For Wiener deblurring, we normalize the blur kernel to sum up to 1 and set $\epsilon = 0.037$. We sufficiently pad the blurred image before deconvolution, and then crop the recovered image.

Measured PSF. Figure 11 showcases the PSFs measured with our prototype for different display layouts. For each display layout, we focus the camera on a white light LED that is placed far away from the camera in a dark room and capture an exposure stack. We fuse each exposure stack into an HDR image, and crop a patch around the brightest point as PSF. Specifically, we crop a patch of 300×300 pixels for a display that has 150 DPI, and 400×400 pixels for a display that has 300 DPI. These PSFs are used both in Wiener deconvolution as well as to train DNNs for deblurring.

Capture settings. We fix ISO to be the smallest value 50 and use an exposure time of $1/125$ s for most outdoor scenes and $1/8$ s for indoor scenes.

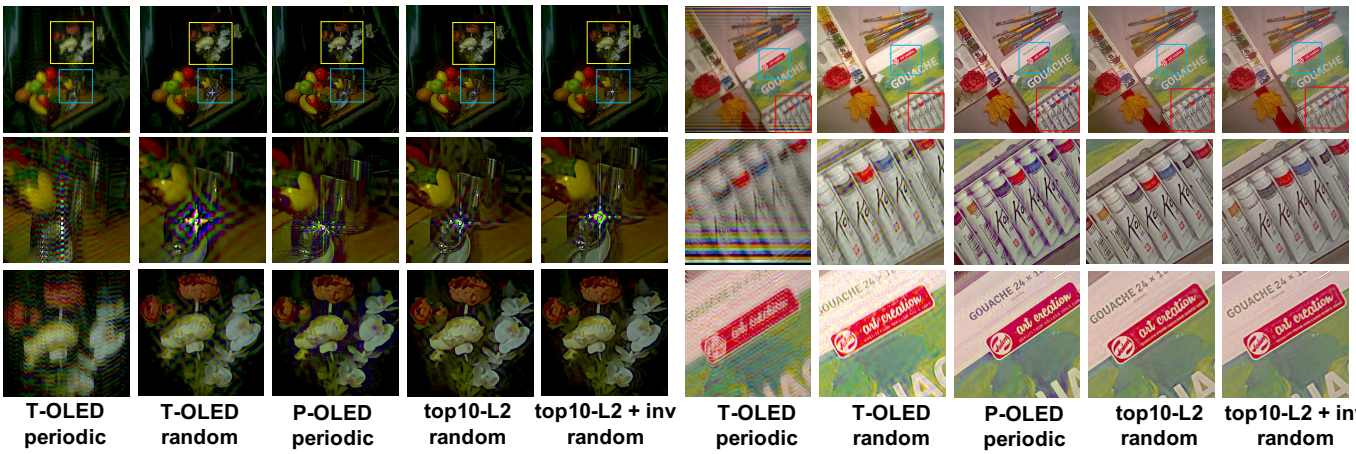


Fig. 12. **Indoor scenes captured by our lab prototype under five different displays.** All display masks have a pixel density of 300 DPI. We show Wiener deconvolved results. Indoor scenes are close to the camera.

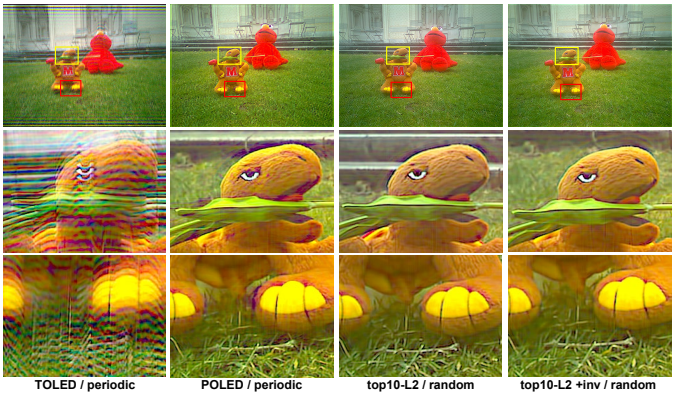


Fig. 13. **Outdoor scenes captured by our lab prototype under five different displays.** All display masks have a pixel density of 300 DPI. Outdoor scenes are relative far-away from the camera and can better satisfy the infinity assumptions.

Results for 300 DPI displays. We show results captured under 300 DPI display patterns for two indoor scenes in Figure 12 and one outdoor scene in Figure 13. Randomly tiled T-OLED has significantly better results than a conventional T-OLED layout. P-OLED and the optimized patterns yield relatively good performance in recovering the details such as texts and edges. However, P-OLED results contain more ringing artefacts; for example, around the toy's feet in Figure 13, purple halos around the flowers and ghosting around the texts on the painting tubes in Figure 12. All methods produce artefacts at specular regions on the spoon; this is a consequence of the non-linearity induced by saturation that violates the linear blur model. We also provide results for the outdoor scene in Figure 13 under 150 DPI displays in the supplementary to characterize performance of the system under lower DPI.

Comparison of deblurring methods. When presented with a large blur kernel, cropping introduced by the sensor has a nontrivial effect on deblurring. We show two additional deblurring methods that implicitly and explicitly handle the boundary issues — a deep neural network and deblurring with an iterative solver — on degraded images captured using our prototype.

For the neural network, which we denote as UNet-RDB, we use the same network structure as in [15], where each layer in the downsampling subnet consists of two Residual Dense Blocks (RDB) [16] and a 2D convolution layer, and each decoding layer has two RDBs and a transpose convolution layer. Since the network is specified to the blur kernel, we train for two networks, one each for T-OLED with periodic tiling and L2+inv with random tiling. For each network, we construct 600 training image pairs and 30 validation pairs using images in HDR+ dataset [33]. We first demosaic RAW images in HDR+ [33] to serve as ground-truth images, and then blur the ground-truth images using captured PSFs and add noise to them. We randomly crop 256×256 patches and use a batch size of 10 in each iteration. All networks are trained for 1000 epochs with a learning rate 1e-3 at the beginning and scaled by 0.1 every 250 epochs. We use Adam optimizer with $\beta_1 = 0.9, \beta_2 = 0.999$.

As we observe in Figure 14, for the T-OLED pattern, the UNet-RDB significantly improves the reconstruction quality, recovering finer details with fewer ringing artefacts as compared to Wiener deconvolution. The improvements for the L2+inv pattern are less subtle, due to the inherent robustness of the PSF; there is some noise suppression, but the network also introduces some artefacts in the process.

For the iterative solver, we model the unknown sharp image to be larger than the known blurred image, and as a convolution of the sharp image with *valid* boundary condition in MATLAB emulating convolution+sensor cropping. In Figure 15, we deblur the teaser images using a linear solver with Tikhonov prior on the image gradients. Compared to T-OLED and P-OLED, the proposed display layout has fewer ringing artefacts along all the edges.

Teaser. Figure 1 shows results on an outdoor scene for the T-OLED, P-OLED, and the Top-10 L2+Inv mask, all three with 300 DPI, and deblurred with Wiener deconvolution. The large spread of the blur in T-OLED along the x-axis leads to severe artefacts. These artefacts are less severe in P-OLED patterns. In contrast, the robustness enabled by our technique results in remarkably better results.

Fig. 14. **Comparison of deblurring methods.** We compare Wiener deconvolution and UNet-RDB on a selfie captured under conventional T-OLED and our optimized display. We use display masks of 150 DPI.

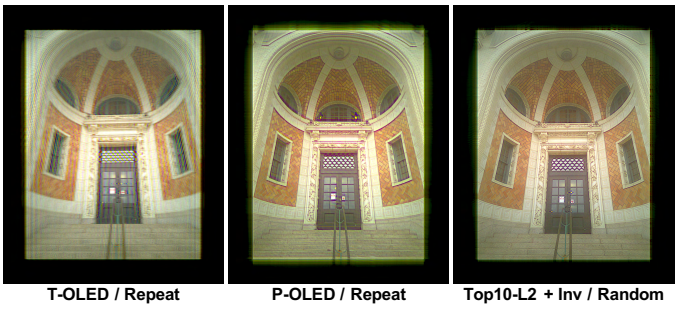


Fig. 15. **Deblurring with TV prior.** We recover sharp images by optimizing least square with TV prior using vanilla linear solver with *valid* boundary condition. Please zoom in to see the details.

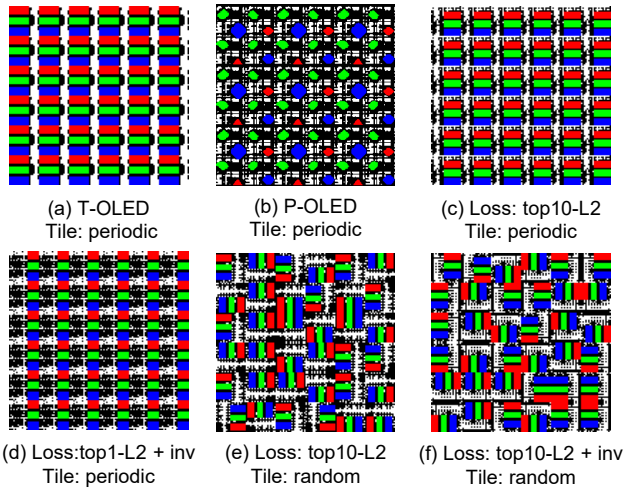


Fig. 16. **RGB subpixel placement for different display layouts.** (a-b) show two typical RGB subpixel placement for OLED screen, and (c-f) show examples of RGB subpixel placement in out optimized display patterns. White region represents pixel openings and red, gree, blue represent regions for R, G, B subpixels.

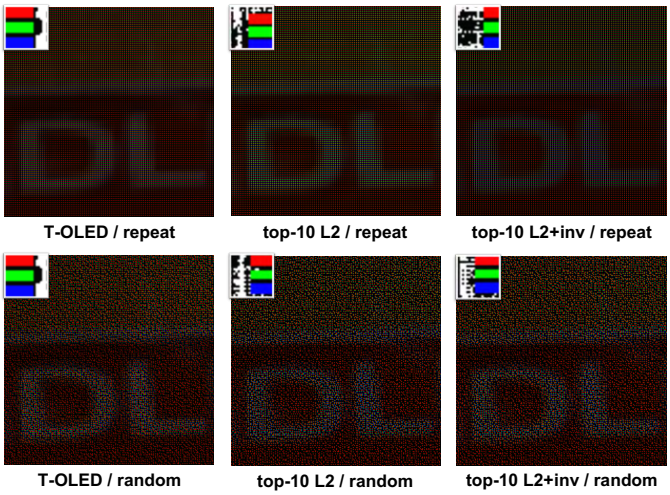


Fig. 17. **Rendering an image using different display layouts.** This is a high-resolution rendering of the displays and each image corresponds to a display size of 8.4mm. To view it at the correct size as it would appear on a display with 300 DPI, please use 25% zoom.

7 Conclusion and Discussions

This paper shows that photographs obtained using under-panel cameras can be improved via careful design of the openings in the display through which the camera observes the scene. We show that introducing non-periodic pixel tilings as well as optimizing the mask openings at each pixel improves the invertibility of the diffractive blur introduced by the display; so much so that, even simple deblurring techniques like Wiener deconvolution can be successful. This indicates that designing the display layout is a promising approach for making under-panel imaging practical.

OLED placement over optimized patterns. It is critical that any change in the display layout accommodate an LED array at the desired resolution, in terms of DPI. We show the RGB subpixel placement for T-OLED and P-OLED displays as well as potential subpixel placement for the proposed display patterns in Figure 16. In all cases, the RGB pixels have the same footprint, in terms of area, although with an OLED placement that is no longer uniform. This non-uniform placement does run the risk of displaying content that appears aliased; however, we contend that this is minimal when operating with high-resolution displays. To validate this, we simulate images on the different display layouts in Figure 17. The displays are at 300 DPI and correspond to a square region with a width of 8.4mm, thereby emulating the area immediately in front of the under-panel camera. The first and the second row show T-OLED and two optimized displays, with periodic tiling and random tiling respectively. We observe that compared to the conventional periodic tiling, random tiled pixels yield reasonable display performance. Although random tiling introduces artefacts, it has a negligible visualization effect at these high

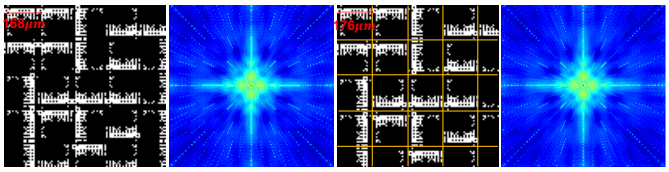


Fig. 18. **Accomodation of opaque wiring.** The left two figures show the proposed Top10 L2+Inv random display and its PSF, and the right two figures show the same pattern with horizontal and vertical space of 8μm added around each pixel.

resolutions. However, the randomness in OLED sub-pixel placement is likely to introduce challenges in manufacturing the display panel as well as designing the wiring pattern for data, control, and power, which we discuss next.

Accommodating power/control wiring. Layout for the wiring required to power and control the OLEDs come in two forms: transparent and opaque [34]. Transparent wiring can be overlaid under the color pixels and usually has a width of half pixels. Given its transparency, it has a negligible effect on our system modeling. Opaque wiring is narrower, taking a width of around 8μm. The simplest approach to adapt opaque wiring to our display design is to add horizontal and vertical space of 8μm around each pixel. As shown in Figure 18 , we add an additional 8μm spacing around each display pixel, whose size is 168μm and it has a negligible effect on the PSF. In all, given the maturity of fabrication technology, we believe that handling the randomized layout is an engineering challenge that can, in principle, be surmounted.

Depth dependence. All of the results in the paper are using blur PSFs measured under the assumption of the scene at infinity. For scenes with points close to the camera, there is a possibility that the PSF at infinity has a significant mismatch to that from a finite distance. To quantify this, we measure the PSF for our L2+Inv optimized mask when we focus on a point light source placed at different depths. This is shown in Figure 19. The observed blur kernel is (near) constant over the depth range that our prototype is capable of focusing on; we provide a detailed theoretical justification for this in the supplemental material. The net result is that we can successfully deblur an in-focus scene immaterial of the depth as seen in Figure 19, using the measured blur kernel for a scene at infinity. We also show that the defocus blur (last row of Figure 19) is stable when deblurred (inset), indicating a gentle bokeh on the out-of-focus regions. These observations are consistent with the deblurred textures in Figure 13, for both the in-focus and out-of-focus regions.

Spatial dependence of the blur PSF. The results shown in the paper assume that the blur kernel is spatially invariant and that the degraded image can be modeled as a convolution between sharp image and a single PSF. In reality, the spacing between the camera lens and the display panel causes the blur PSF to be spatially varying. Further, non-idealities in the lens introduces other aberations including the Pincushion distortion seen in the PSFs shown in Figure 20. We show residual blurs (insets) of all PSFs when deblurred by the center PSF.

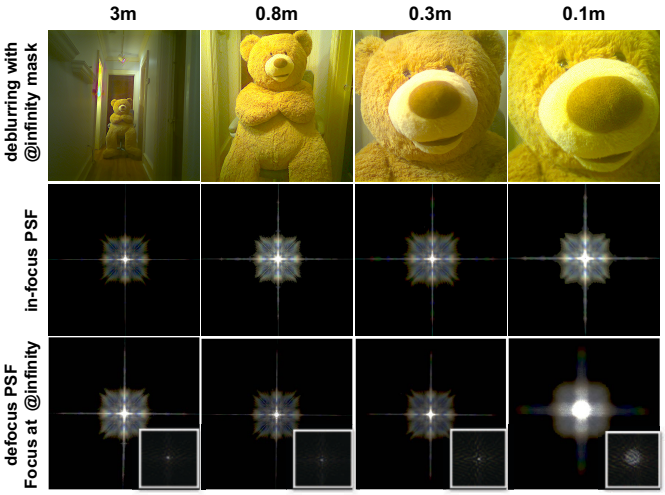


Fig. 19. **Deblurring at different depths using the Top10-L2+Inv pattern.** (Top row) Deblurred photos for in-focus scene at different depths. In all cases, the deblurring was performed with the PSF corresponding to scene-at-infinity. (middle row) PSF of a point light source at different depths, with camera focused on the light source. (bottom) PSF of a point light source with camera focused at infinity. (bottom-inset) Residual blur after deblurring the defocus blur.

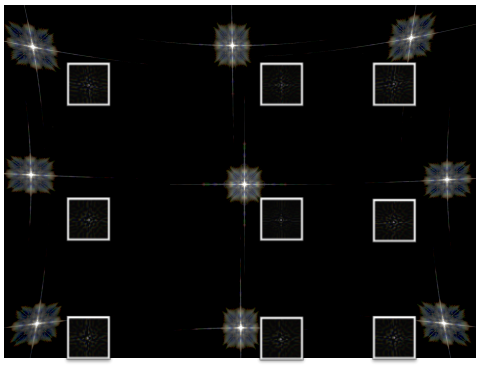


Fig. 20. **Spatial variation of the blur kernel of the Top10-L2+Inv pattern.** We capture PSFs that appear on all corners and near edges. Insets show residuals after deblurring with the center PSF.

Handling saturation. Non-linearities in the imaging pipeline can create a significant model mismatch to the linear model commonly assumed in deblurring approaches. For example, specular highlights in Figure 12 leads to jarring artefacts in the restored photographs. We can handle such scenes either by incorporating such artefacts in the training dataset for the DNN model. Alternatively, capturing an HDR image by exposure bracketing to revert back to a linear model helps, as seen in Figure 21.

ACKNOWLEDGMENTS

This work was supported by Global Research Outreach program of Samsung Advanced Institute of Technology and the NSF CAREER award CCF-1652569. The authors thank Drs. Eunhee Kang and Hyong-Euk Lee for introducing them to this problem and valuable discussions on display design.

Fig. 21. **Saturated regions.** Saturated pixels break the linearity of the imaging model and leads to artifacts in the deblurred photograph. We alleviate these by using HDR photography to obtain a linear blur model.

REFERENCES

[1] T. Tsujimura, *OLED display fundamentals and applications*. John Wiley & Sons, 2017.

[2] A. Yang and A. C. Sankaranarayanan, "Design display pixel layouts for under-panel cameras: Software," https://github.com/Image-Science-Lab-cmu/UPC_ICCP21_Code, 2021.

[3] B. Masia, G. Wetzstein, P. Didyk, and D. Gutierrez, "A survey on computational displays: Pushing the boundaries of optics, computation, and perception," *Computers & Graphics*, vol. 37, no. 8, pp. 1012–1038, 2013.

[4] A. D. Wilson, "TouchLight: An imaging touch screen and display for gesture-based interaction," in *International Conference on Multimodal interfaces*, 2004.

[5] H. Benko and A. D. Wilson, "DepthTouch: Using depthsensing camera to enable freehand interactions on and above the interactive surface," in *IEEE Workshop on Tabletops and Interactive Surfaces*, 2009.

[6] M. Hirsch, D. Lanman, H. Holtzman, and R. Raskar, "BiDi screen: A thin, depth-sensing lcd for 3d interaction using light fields," *ACM Trans. Graphics*, vol. 28, no. 5, pp. 1–9, 2009.

[7] C. J. Cheng, T. C. Huang, W. T. Lin, C. C. Hsieh, P. Y. Chen, P. Lu, and H. Y. Lin, "Evaluation of diffraction induced background image quality degradation through transparent oled display," in *SID Symp. Digest of Technical Papers*, 2019.

[8] Y. Zhou, D. Ren, N. Emerton, S. Lim, and T. Large, "Image restoration for under-display camera," *arXiv preprint arXiv:2003.04857*, 2020.

[9] N. Emerton, D. Ren, and T. Large, "Image capture through tft arrays," in *SID Symp. Digest of Technical Papers*, 2020.

[10] Z. Zhang, "Image deblurring of camera under display by deep learning," in *SID Symp. Digest of Technical Papers*, 2020.

[11] S. Lim, Y. Zhou, N. Emerton, T. Large, and S. Bathiche, "Image restoration for display-integrated camera," in *SID Symp. Digest of Technical Papers*, 2020.

[12] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Intl. Conf. Medical image computing and computer-assisted intervention*, 2015.

[13] V. Sundar, S. Hegde, D. Kothandaraman, and K. Mitra, "Deep atrous guided filter for image restoration in under display cameras," *arXiv preprint arXiv:2008.06229*, 2020.

[14] D. Puthussery, M. Kuriakose, J. C V *et al.*, "Transform domain pyramidal dilated convolution networks for restoration of under display camera images," *arXiv preprint arXiv:2009.09393*, 2020.

[15] Y. Zhou, M. Kwan, K. Tolentino, N. Emerton, S. Lim, T. Large, L. Fu, Z. Pan, B. Li, Q. Yang *et al.*, "Udc 2020 challenge on image restoration of under-display camera: Methods and results," *arXiv preprint arXiv:2008.07742*, 2020.

[16] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu, "Residual dense network for image super-resolution," in *CVPR*, 2018.

[17] E. E. Fenimore and T. M. Cannon, "Coded aperture imaging with uniformly redundant arrays," *Applied Optics*, vol. 17, no. 3, pp. 337–347, Feb 1978.

[18] A. Levin, R. Fergus, F. Durand, and W. T. Freeman, "Image and depth from a conventional camera with a coded aperture," *ACM Trans. Graphics*, vol. 26, no. 3, pp. 70–es, 2007.

[19] C. Zhou and S. Nayar, "What are good apertures for defocus deblurring?" in *ICCP*, 2009.

[20] C. Zhou, S. Lin, and S. Nayar, "Coded aperture pairs for depth from defocus," in *ICCV*, 2009.

[21] A. Veeraraghavan, R. Raskar, A. Agrawal, A. Mohan, and J. Tumblin, "Dappled photography: Mask enhanced cameras for heterodyned light fields and coded aperture refocusing," *ACM Trans. Graphics*, vol. 26, no. 3, p. 69, 2007.

[22] E. R. Dowski and W. T. Cathey, "Extended depth of field through wave-front coding," *Applied optics*, vol. 34, no. 11, pp. 1859–1866, 1995.

[23] S. R. P. Pavani and R. Piestun, "Three dimensional tracking of fluorescent microparticles using a photon-limited double-helix response system," *Optics express*, vol. 16, no. 26, pp. 22 048–22 057, 2008.

[24] Y. Wu, V. Boominathan, H. Chen, A. Sankaranarayanan, and A. Veeraraghavan, "PhaseCam3D—learning phase masks for passive single view depth estimation," in *ICCP*, 2019.

[25] J. W. Goodman, *Introduction to Fourier optics*. Roberts & Co., 2005.

[26] Q. Sun, E. Tseng, Q. Fu, W. Heidrich, and F. Heide, "Learning rank-1 diffractive optics for single-shot high dynamic range imaging," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 1386–1396.

[27] C. A. Metzler, H. Ikoma, Y. Peng, and G. Wetzstein, "Deep optics for single-shot high-dynamic-range imaging," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020.

[28] J. Chang and G. Wetzstein, "Deep optics for monocular depth estimation and 3d object detection," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019.

[29] K. Mitra, O. S. Cossairt, and A. Veeraraghavan, "A framework for analysis of computational imaging systems: role of signal prior, sensor noise and multiplexing," *IEEE transactions on pattern analysis and machine intelligence*, vol. 36, no. 10, pp. 1909–1921, 2014.

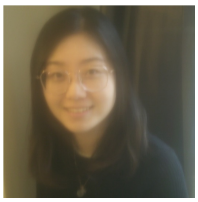
[30] A. Shrivastava, A. Gupta, and R. Girshick, "Training region-based object detectors with online hard example mining," in *CVPR*, 2016.

[31] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-d transform-domain collaborative filtering," *IEEE Trans. image processing*, vol. 16, no. 8, pp. 2080–2095, 2007.

[32] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Trans. image processing*, vol. 13, no. 4, pp. 600–612, 2004.

[33] S. W. Hasinoff, D. Sharlet, R. Geiss, A. Adams, J. T. Barron, F. Kainz, J. Chen, and M. Levoy, "Burst photography for high dynamic range and low-light imaging on mobile cameras," *ACM Trans on Graphics*, no. 6, 2016.

[34] Z. Wang, Y. Chang, Q. Wang, Y. Zhang, J. Qiu, and M. Helander, "Self-assembled cathode patterning in AMOLED for under-display camera," in *SID Symposium Digest of Technical Papers*, 2020.



Anqi Yang is a PhD student in the ECE department at CMU. Her research interests are broadly in computational photography, especially relating to problems in mobile photography. Anqi did her Bacherlors in Engineering in Tongji University, and Masters in Robotics at Carnegie Mellon University. She is the recipient of the Google Anita Borg Memorial Scholarship in 2016, one among nine students who received it in China.

Aswin C. Sankaranarayanan is an associate professor in the ECE department at CMU, where he is the PI of the Image Science Lab. His research interests are broadly in compressive sensing, computational photography, signal processing and machine vision. His doctoral research was in the University of Maryland where his dissertation won the distinguished dissertation award from the ECE department in 2009. Aswin is the recipient of the CVPR 2019 best paper award, the CIT Dean's Early Career Fellowship, the Spira Teaching award, the NSF CAREER award, the Eta Kappa Nu (CMU Chapter) Excellence in Teaching award, and the Herschel Rich Invention award from Rice University.