Image Restoration for Display-Integrated Camera

Sehoon Lim*, Yuqian Zhou**, Neil Emerton*, Tim Large*, Steven Bathiche*
***Microsoft Applied Sciences, Redmond WA**
**** IFP Group, UIUC**

Abstract
Under-display camera is of great interest in the display industry potentially eliminating the display bezel and camera notch/hole in mobile devices. However, display panels cause complex signal modulation in the camera aperture which results in obscuration, attenuation and diffraction of the incident light. We propose a learning-based image restoration approach to enable a camera to operate underneath the display without affecting the display contrast and color gamut.

Author Keywords
Under-display camera; display panel; organic light-emitting diode (OLED); diffraction; point-spread function (PSF); signal-to-noise ratio (SNR); peak signal-to-noise ratio (PSNR); modulation transfer function (MTF); image restoration; image denoising; image deblurring; deconvolution; neural network (NN); convolutional neural network (CNN); deep neural network (DNN).

1. Introduction
There has long been an interest in locating imaging systems behind or underneath displays. Transparent displays based-on LCD and OLED have been released, albeit with low resolution and reduced display contrast and color gamut [1, 2, 3]. Recently a smartphone vendor showed an under-display selfie camera with a customized low-resolution and transparent display region centered on the camera location [4]. Optical fingerprint sensors under a smartphone's OLED display have been demonstrated but these are not capable of high-resolution, color imaging [5]. We present the combination of high resolution and color imaging through a high-quality display.

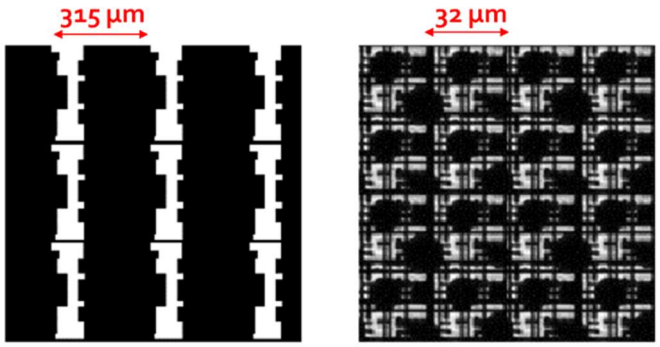
Placing a camera under a display conflicts with the needs of high-quality camera imaging which conventionally requires a clear aperture to receive enough uninterrupted light from the scene. The display panel in front of the camera prevents fulfillment of the imaging requirements by modulating the incident light due to the display's 3D structure. The display panel is typically composed of stacked optical layers such as polarizers, pixel structures, and a substrate. The pixel structure is purely opaque in metal TFTs whose lateral design determines the display's pixel layout, resolution, light attenuation and resulting diffraction for camera imaging. Ideally a favorable display structure could be designed; however, the technology constraints in the display design limit what can be achieved. Furthermore, the display and camera industries are so separated that we were motivated to investigate these imaging problems and construct solutions with existing display panels and cameras.

Computationally, image deconvolution is well-established to reconstruct the original object from the blurred image [6, 7]. Deconvolution is the inverse process of convolution and recovers the original signal from the point-spread-function (PSF)-convolved image. The fidelity of the deconvolution process is dependent on the space-invariance of the PSF over the image field-of-view (FOV) and on a low condition number for the inverse of the PSF [8]. For strongly non-delta-function-like PSFs such as those encountered when imaging through a display, the value of condition number can be large. For such PSFs an additional denoising step may be essential.

In contrast, learning-based methods are in essence a complex fitting process which is decoupled from the above mathematical formalism. The high numerical flexibility in these methods also permits that the pre-defined PSF is space-dependent. In this paper, we propose a UNet architecture which is a U-shaped neural network composed of 3×3 convolutional layers and activation functions in the contraction and expansion paths [9]. The U-Net model preserves the local feature in the contraction path and transfers it to the expansion path. The learnable parameters are designed by the model structure and the parameter values are determined by data training.

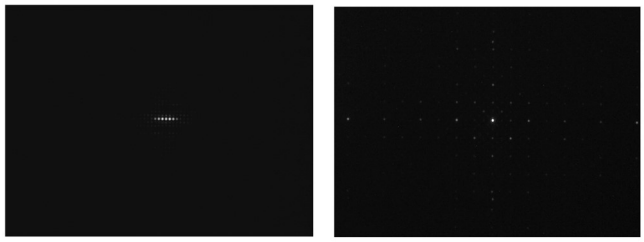
This paper covers the full scope of problem definition, system characterization, and learning-based image restoration. The major factors of image degradation are addressed by the 3D structure of display panel. The performance of our proposed method is characterized by measuring the modulation transfer function (MTF) using a slanted-edge test and the signal-to-noise ratio (SNR), enabling us to define a system resolution and noise-power budget. The system budget clearly shows what we can achieve now, and what may be achievable in the future. The learning-based method is discussed in terms of the model design and the data collection. Finally, we demonstrated the effectiveness of this approach by recovering complex test images.

2. System Characteristics
Two types of display panels are discussed for our case study as shown in Figure 1. One is a transparent OLED (tOLED) for a large-scale 4K display with a pixel pitch of 315 micron; the transmittance is 18.5% with 20% open area on a clear substrate. The other is a Pentile OLED (pOLED) for mobile device. The Pentile structure is evidently more complex than the stripe structure of tOLED. The transmittance is only 3% with 23% open area and there is attenuation and color shift due to the yellow polyimide substrate.



(a) tOLED (b) pOLED

Figure 1. Two samples of display panels: (a) tOLED and (b) pOLED



(a) tOLED (b) pOLED

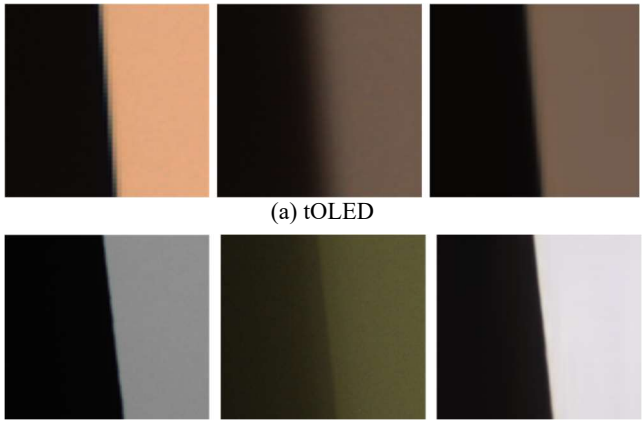
Figure 2. Point-spread functions through the two display panels: (a) tOLED and (b) pOLED

Image degradation mainly arises from contrast reduction resulting from diffraction from the pixel pattern, and signal attenuation from the stacked layers in the display panel. Figure 2 shows PSFs measured through tOLED and pOLED samples using a HeNe laser at 633nm. The rectangular slits of the tOLED screen act like an amplitude diffraction grating and produce six strong side lobes only a few pixels away from the main lobe. However, the complex structure of the pOLED screen diffracts light into many sparse side lobes, each of which is relatively weaker than those of the tOLED.

To evaluate the image degradation and restoration, we used the slant edge method that measures the spatial frequency response as an approximation of the modulation transfer function [10, 11, 12]. In the test, a tilted square pattern is captured through the display samples and the raw image is recovered by our image restoration method (Figure 3). In the tOLED case, light attenuation occurs due to the 18.5% sample transmission and the edge sharpness falls strongly owing to the diffracted PSF. In the pOLED case, light attenuation and color shift occur due to the 3% transmission and yellow polyimide substrate. Although the noise impact is very strong in the pOLED, the overall subjective loss in sharpness is not very bad compared to that of the tOLED. In both cases, the recovered images are subjectively comparable to the original image.

To quantitatively analyze the image data, we used a linear integral over the frequency response, although other single-figure image quality metrics could be used.

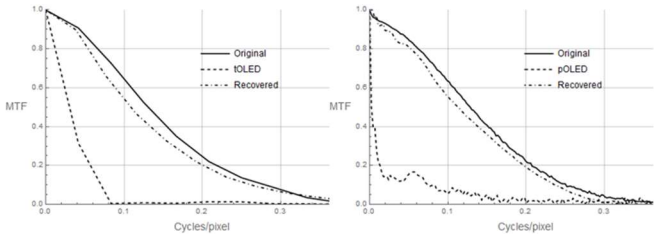
$$\text{Linear Integral} = \int \text{MTF(f) df}$$



(a) tOLED

(b) pOLED

Figure 3. Slant edge test for (a) tOLED and (b) pOLED: original, raw, and recovered images



(a) tOLED (b) pOLED

Figure 4. MTFs for (a) tOLED and (b) pOLED

The tOLED test resulted in integrated contrasts for the original (0.147), raw (0.043), and recovered (0.133) images. In the same manner, the pOLED test resulted in integrated contrasts for the original (0.138), raw (0.024), and recovered (0.124) images. The image restoration recovered 90% contrasts for both the tOLED and pOLED samples in Figure 4.

The system performance is summarized by tabulating the characteristics of the original, raw, and recovered images in terms of MTF and signal-to-noise ratio (SNR) as shown in Table 1 and 2. Note that the Neural Network (NN) gain indicates the image improvement produced by the NN in the contexts. The MTF table shows the NN gain from the blurry raw image to the sharp recovery. The tOLED and pOLED show the NN gains of 3 and 5.29 resulting in 90% contrasts in the recovery. The SNR table shows how the NN gain recovers the noisy image. The tOLED and pOLED show the NN gains of 2 dB and 10 dB resulting in -5.0 dB and -5.2 dB SNR loss. The SNR was obtained by using a selected region of 100 by 200 pixels in the image data. Therefore, the learning-based method mainly focused on deblurring to improve the contrast and denoising to improve the SNR in the samples. This performance trend is determined by the data training dependent with the display samples. The MTF performance is fixed by both the panel design and the NN gain, however, the SNR performance could be improved by increasing the number of measurements.

Table 1. MTF budget for tOLED and pOLED

Camera under tOLED		Camera under pOLED	
	Fraction		Fraction
Contrast	0.3	Contrast	0.17
NN gain	3	NN gain	5.29
Total Loss	0.9	Total Loss	0.9

Table 2. SNR budget for tOLED and pOLED

Camera under tOLED			Camera under pOLED		
	Fraction	dB		Fraction	dB
Transmission	0.2	-7.0	Transmission	0.03	-15.2
NN gain	1.58	2.0	NN gain	10	10.0
Total Loss	0.32	-5.0	Total Loss	0.3	-5.2

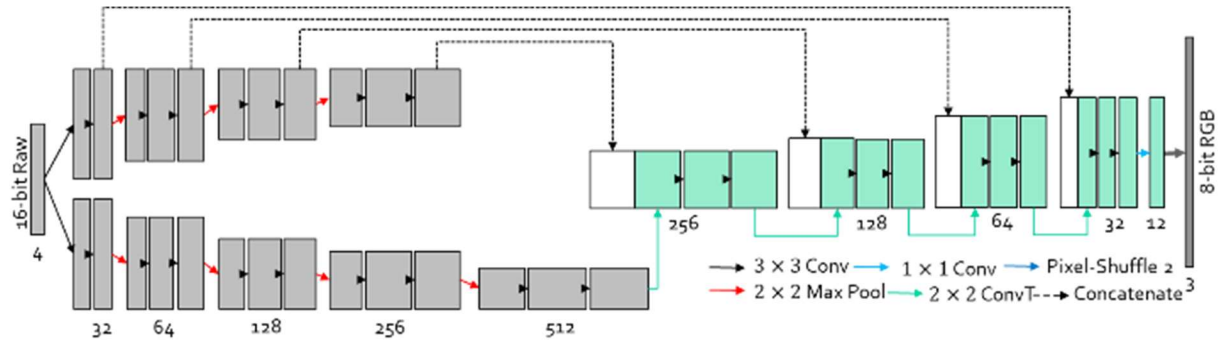


Figure 5. Network structure for Unet

3. Learning-based Image Restoration

The degraded measurement $\hat{y}$ is formulated by the convolution of the original image x and the PSF adding the noise n. The PSF represents the blur kernel resulted from diffraction. The image reconstruction $\hat{x}$ is modeled by solving the maximum a posteriori (MAP) problem composed of the least squares term and the regularization term [13].

$$\hat{x} = \text{argmin}_x \frac{1}{2\sigma^2}\|\hat{y} - y\|^2 + \lambda\Phi(x),$$

Where σ is the noise level and $\Phi(x)$ is the regularization term. The objective function of the L-1 loss $\mathcal{L}_1(\Theta)$ is applied to train the model for image reconstruction. The reconstruction $\hat{\mathcal{F}}(\mathbb{y}_i\,;\,\Theta)$ is estimated by the observation of degraded image $\mathbb{y}_i$ and the learnable network parameters Θ, and it is compared to the original image $\mathbb{x}_i$ to calculate the loss. The network parameters Θ are defined by the model structure and they are learned within a training batch of images.

$$\mathcal{L}_1(\Theta) = \frac{1}{N}\sum_{i=1}^{N}\left\|\hat{\mathcal{F}}(\mathbb{y}_i\,;\,\Theta) - \mathbb{x}_i\right\|_1,$$

Where N is the total number of images inside a training batch. A UNet structure was used to train the image restoration model which splits the encoder into two sub-encoders. One sub-encoder stores residual details for the decoding process and the other learns contents from encoding the degraded image. The proposed model takes a 4-channel raw 16-bit image (y) and returns a RGB 8-bit recovery (x) as shown in Figure 5.

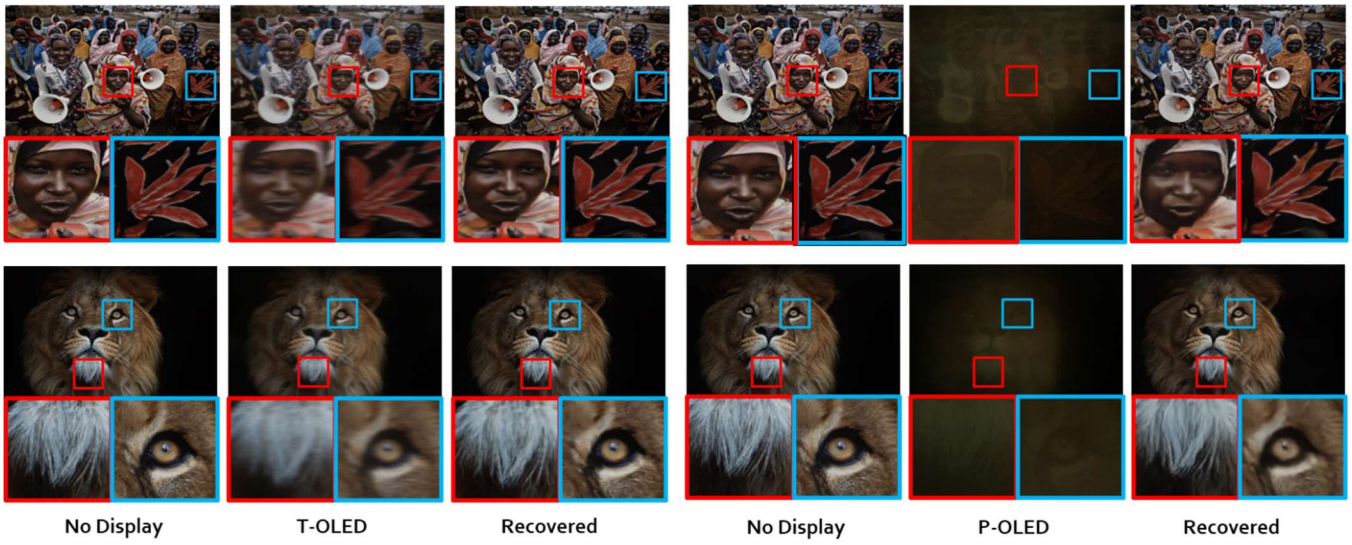
A display-camera imaging system was designed to collect a set of training dataset degraded by display samples. A 4K LCD display showed the training dataset images sequentially. A 12 MP Point-Grey camera with on-camera image binning producing 2 MP images was used to image the display at a distance of 30 cm. The camera lens of aperture F/1.8 was focused on the 4K LCD's content. The camera operates at 8 FPS, 125 ms shutter speed, and data is output using raw 16-bit image format. The low frame rate and long exposure time resulted from the light attenuation via the tOLED and pOLED display samples. Furthermore, the camera gain was set to 6 dB for tOLED sample and to 25 dB for pOLED to compensate the different level of light attenuation. The same camera conditions are also used in the tOLED and pOLED reconstruction steps. The overall setup for data collection was enclosed by a black box to avoid any impact from ambient illumination.

To adjust the monitor gamma, 2.2 Transform is set during the data collection. The data captures with and without the display sample are well-aligned to each other, typically within one or two pixels; however, any small image shifts caused by the display sample are adjusted by off-line image registration. Unet requires the power of 2 training size and only central region of (1024, 2048) pixels is cropped from original 2 MP of (1040, 2048) pixels for NN training. The collected data image set was split as 200 training, 40 validation, and 60 testing. This number of images is relatively small, but the images are themselves relatively large compared to the number of free parameters in the network. Increasing the number of images did not result in discernible improvement in image quality of the final trained images. We trained the model using the Adam optimizer with a learning rate of 1e-4 and a decay factor of 0.5 after 200 epochs. The training stopped at epoch 400 and the best validation performance was selected.

The results of image restoration trained on the pair-wise data are shown in Figure 6. The recovered images are sharper and less noisy for both the tOLED and pOLED cases although some of image features are lost from the original images. The recovery differently improved the sharpness and the noise for the two samples. The quantitative results are also reported by the peak signal-to-noise ratio (PSNR) which is the difference between the recovered image and the original image. In the tOLED the PSNR increased from the raw image of 28.8 dB to the recovered image of 36.7 dB by the gain of 7.9 dB. In the pOLED the PSNR increased from the raw image of 15.4 dB to the recovered image of 30.5 dB by the gain of 15.1 dB. The tOLED has the higher PSNR value in the recovered images but the pOLED has the higher PSNR gain.

4. Conclusion and Discussion

The performance of an under-display camera is optically defined and characterized by MTF and SNR in two types of display samples. A pairwise data collection method between training data and display-degraded data is implemented using a high-resolution LCD display. A learning-based model is designed and trained to deal with both diffraction blur and noise in the image degradation. However, the inherent low SNR issue is not fully solved by this implementation of the learning-based method. In conclusion, we hope that this paper motivates the display industry to consider the needs of under-display cameras when designing future pixel structure.



(a) tOLED (b) pOLED

Figure 6. Image restoration for (a) tOLED and (b) pOLED.

5. Acknowledgements

The authors are grateful to Dr. SooYoung Yoon and Dr. JoonYoung Yang in LG Display for providing a sample pixel layout and for helpful discussions on this project.

6. References

1. Planar Systems Corp. https://www.planar.com/products/transparent-oled-displays/
2. YDEA Group. https://www.ydea.group/transparent-led-display/
3. Crystal Display Systems Ltd. http://crystal-display.com/category-transparent/
4. The Verge. https://www.theverge.com/2019/6/26/18759380/under-display-selfie-camera-first-oppo-announcement; online magazine.
5. Synaptics Inc. https://www.synaptics.com/products/biometrics
6. Candes E, Romberg J, and Tao T. Robust uncertainty principles: Exact recovery from highly incomplete Fourier information; IEEE Transactions on Information Theory; February 2006.
7. Li C. Compressive Sensing for 3D Data Processing Tasks: Applications, Models and Algorithms; PhD Thesis; Rice University; 2011.
8. Heath M. Scientific Computing: An introductory survey; second edition; 2002.
9. Ronneberger O, Fischer P, and Brox T. U_Net: Convolutional Networks for Biomedical Image Segmentation; Computer Vision and Pattern Recognition; May 2015.
10. CPIQ P. Standard for camera phone image quality. Institute of Electrical and Electronics Engineers (IEEE); 2015.
11. ISO/TC42/WG18. Resolution and spatial frequency response. International Organization for Standardization (ISO); 2000.
12. ISO/TC42/WG18. Resolution and spatial frequency response. International Organization for Standardization (ISO); 2014.
13. Zhang K. Zuo W. Zhang L. Learning a Single Convolutional Super-Resolution Network for Multiple Degradations; CVPR; 2018