International Journal of Multidisciplinary Research and Growth Evaluation

International Journal of Multidisciplinary Research and Growth Evaluation
ISSN: 2582-7138
Received: 27-04-2021; Accepted: 13-05-2021
www.allmultidisciplinaryjournal.com
Volume 2; Issue 3; May-June 2021; Page No. 185-189

Image restoration for under-display camera based on deep learning

Duc D Nhu [1], Giao N Pham [2], Binh A Nguyen [3], Manh Hoanh [4], Tung V Nguyen [5]
[1] Department of Multimedia Engineering, Dongguk University Seoul, South Korea
[2] Department of Computing Fundamentals, FPT University, Hanoi, Vietnam
[3, 4, 5] ICT Department, FPT University, Hanoi, Vietnam

Corresponding Author: **Duc D Nhu**

Abstract
The new trend of full-screen devices encourages us to position a camera behind a screen. Removing the bezel and centralizing the camera under the screen brings larger display-to-body ratio and enhances eye contact in video chat, but also causes image degradation. In this paper, we propose new network architecture that can enhance the image captured by Under-Display Camera. Our research focuses on the image from the camera under T-OLED display.

Keywords: Under-Display Camera, Image restoration, Image deburring

1. Introduction

Under-display Camera (UDC) [1] is a new imaging system that mounts a display screen on top of a traditional digital camera lens. Such a system has mainly two advantages. First, it follows a new product trend of full-screen devices with a larger screen-to-body ratio, which can provide better user perceptive and intelligent experience. Without seeing the bezel and extra buttons, users can easily access more functions by directly touching the screen. Second, it provides a better human-computer interaction. By putting the camera in the center of the display, it enhances teleconferencing experiences with perfect gaze tracking, and it is increasingly relevant for larger display devices such as laptops and TVs.

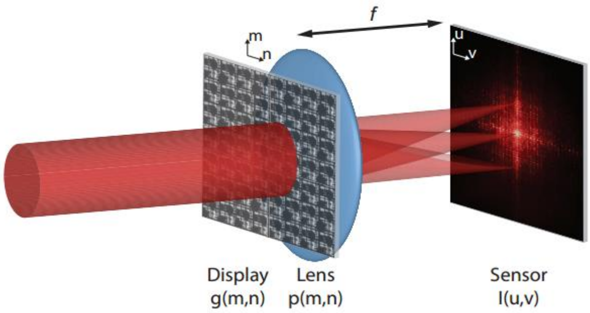

Fig 1: 3D schematics

Unlike pressure or fingerprint sensors that can be more easily integrated into a display, the imaging sensor is relatively hard to maintain its functions after being mounted behind a display. The imaging quality of a camera will be severely degraded due to lower light transmission rates and diffraction effects. As a result, images captured will be noisy and blurry.

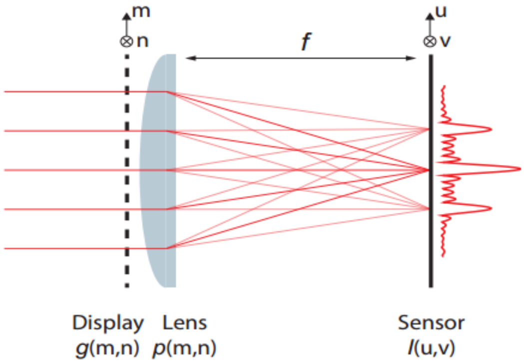

Fig 2: 2D schematics

In this paper, we focus on creating new neural network architecture that restores the quality of an image captured by a camera put under a T-OLED display.

2. Related work

Our research focuses on restoring the image quality of a degraded image, a vast field in image processing, especially for the case of Under-Display Camera (UDC). Zhou *et al.* [1] introduced UDC that is a new imaging system that mounts the display screen on top of a traditional digital camera lens. There are two types of display panels the paper study on T-OLED and P-OLED. They also propose two methods to make data for training and testing; those are Monitor-Camera Imaging System and Realistic Data Synthesis Pipeline. They create an Unet-base architecture for image restoration.

This problem relates to the optic field, specifically the lens less camera system. Sun *et al.* [2] invent a method for snapshot HDR imaging by learning an optical HDR encoding in a single image that maps saturated highlights into neighboring unsaturated areas using a diffractive optical element (DOE). They introduce an optical model and a deep neural net model to synthesize the HDR image from the image captured by the proposed optical model. Khan *et al.* [3] focus on using a deep neural network to increase the quality of the image captured by the lens less Flat-Cam system. There are two main components of their proposed network: Trainable Inversion Layer that create a low-quality image then enhances the image in the next phase, and Perceptual Enhancement Layer uses UNet architecture to enhance the low-quality image from Trainable Inversion Layer to high-quality image.

With the development of the deep neural network for image restoration, there is a lot of works try to inverse the degraded image to a higher quality image that is approximately the original. Kaufman *et al.* [4] divide the problem into two sub-problems; those are specifying the blur kernel of an image by Analysis Network and non-blind de-blurring image using kernel found above by Synthesis Network. Nah *et al.* [5] proposes a method to deal with the problem of dynamic scene blurring: non-uniform blur kernel. I think this kind of problem is quite similar to our problem. Because the focus blurring is depended on the depth of the object to the sensor, then it makes different blur levels on different objects in the scene. Therefore, their architecture tries to analyze the image in multi-scales, then synthesize the final image at the original scale. Kupyn *et al.* propose DeblurGAN [7] and DeblurGANv2 [8], which are end-to-end learned methods for motion de-blurring based on a conditional GAN and the content loss. They achieve state-of-the-art performance both in the structural similarity measure and visual appearance. Especially, DeblurGANv2 using the plug-and-play method for balancing the performance and computational cost, which can easily deploy on a low computational resource like the mobile device. Zhang *et al.* [9] propose a more realistic degradation model than the bi-cubic degradation model for SISR. It considers arbitrary enables to use of existing de-blurring methods for blur kernel estimation and a deep plug-and-play super-resolution framework (DPSR) for solving SISR with the new degradation model. It can handle low-resolution images with arbitrary blur kernels. The proposed DPSR is well-principled as the iterative scheme aims to address the new degradation induced energy function. Lim *et al.* [10] proposed two architectures based on Enhanced residual blocks, which are EDSR for single-scale image super-resolution and MDSR for multi-scale image super-resolution. They modify the residual block due to some observation, and it makes the model get a higher quality image. Base on Lim *et al.*'s [10] work, Zhang *et al.* [11] propose Residual Dense Network (RDN), which built up from Residual Dense Block (RDB). Their architecture utilizes the hierarchical feature maps and achieves the state-of-the-art result on the image super-resolution as well as the de-blurred image. Beside of that, the new loss function is invented for making the model achieves a higher quality image. Johnson *et al.* [6] introduce perceptual loss that the perceptual depending on high level features extracted from a convolutional network. The convolutional neural networks the pre-trained for image classification have already learned to encode the perceptual and semantic information. Image generated by a model trained by perceptual loss is usually more "meaning."

Given an object in scene **x**, the degraded observation **y** can be modeled:

$$\mathbf{y} = (\gamma \mathbf{x}) \otimes \mathbf{k} + \mathbf{n} \quad (1)$$

Where **γ** is the intensity scaling factor under the current gain setting and display type, **k** is the PSF, and **n** is the zero-mean signal-dependent noise. Notice that this is a simple noise model that approximately resembles the combination of shot noise and readout noise of the camera sensor. An ideal PSF shall resemble a delta function, which then forms a perfect image of the scene. However, light greatly spreads out in UDC. For T-OLED, light spreads mostly across the horizontal direction due to its nearly one dimensional structure in the pixel layout, while for P-OLED, light is more equally distributed as the pixel layout is complex. Therefore, images captured by UDC are either blurry (T-OLED) or hazy (P-OLED). Modulation Transfer Function (MTF) is another important metric for an imaging system, as it considers the effect of the finite lens aperture, lens performance, finite pixel size, noise, nonlinearities, quantization (spatial and bit depth), and diffraction in our systems. For T-OLED, contrasts along the horizontal direction are mostly lost in the mid-band frequency due to diffraction. This phenomenon is due to the one-dimensional pixel layout of the T-OLED. Fig. 3 shows severe smearing horizontally when putting T-OLED in front of the camera. While for P-OLED, the MTF is almost identical to that of a display-free camera, except with severe contrast loss. Fortunately, however, nulls have not been observed in any particular frequencies.

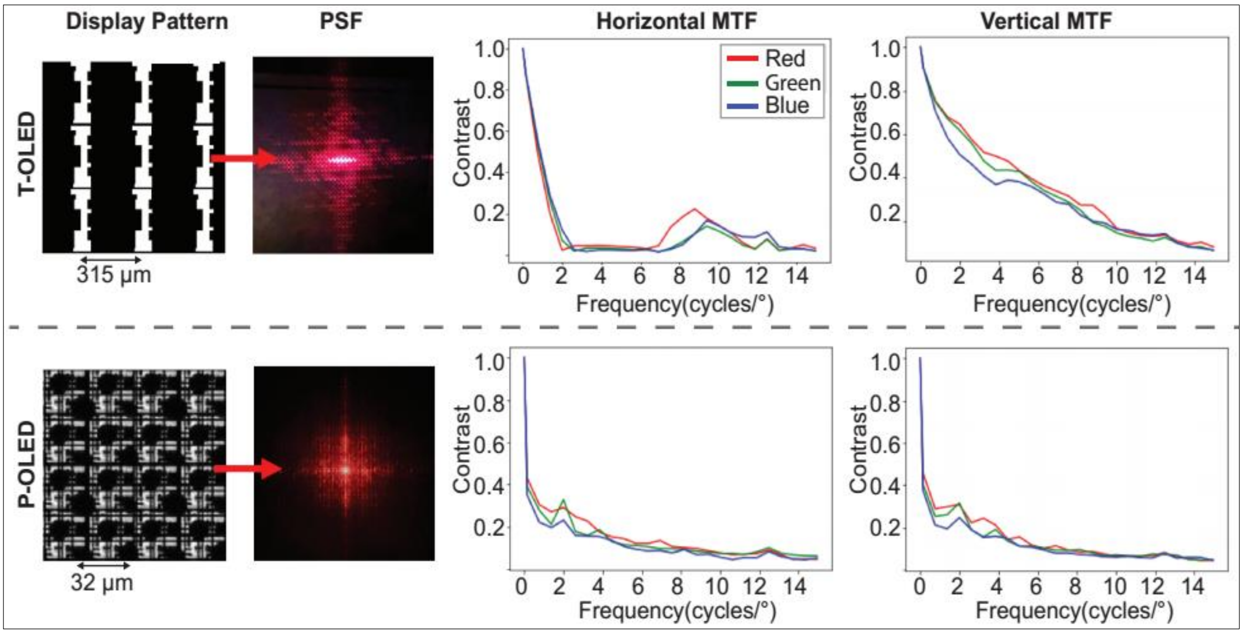

Fig 3: Optics characteristics of UDC

3. Proposed method

3.1 Architecture

In the paper [1], they assumed that the blur kernel is unique for every part of the image. Actually, because of objects with different depths, they will have different levels of focus, so the blur problem here will be the dynamic blur kernel. With the setup of the camera when capturing images in [1], our problem is considered as the combination of de-blurring, de-noising, and color enhancement.

As the analysis above, we need to recover image differently at a different range of frequency. So my architecture (Fig. 4) will deal with varying levels of scale of the image. We design four levels of scale and try to recover image at four scales

- Down sampling x8 level: It deals with the low-frequency component. I use the bi-cubic technique for both down-sampling and up-sampling. Because we don't want any artifact effect (i.e., adding detail to the image) by a learned filter, it could destroy the recovered low-frequency result. The process component at this level is the simplest because it only processes a small part of the image. Make the process simple also help reduce the computational cost as well as avoid overfitting. It has a global skip connection for learning more effectively.
- Down sampling x4 level: It deals with a larger range of frequency than x8 level. Use a result from x8 level as prior information to make it focus on recovering the higher range of frequency. The process component at this level is more complex than x8
- Down sampling x2 level is quite the same with x4 level
- At original size: It deals with recover the detail of the image. It has the most complicated process component because it must resolve the noise problem. While at other scales, the noise is almost omitted (the noise, especially white-noise, is usually on the high-frequency range). Because it must resolve the noise problem. While at other scales, the noise is almost omitted (the noise, especially white-noise, is usually on the high-frequency range).

At the final step, it accumulates the result from all other levels to make the final image. The processing block (Fig. 5) will be build based on the primary building block. That is Residual Dense Block [11] (Fig. 6). The numbers of building blocks on each processing block [1, 2, 3, 4] (Fig.3) are different. In our setup, we set it to [20, 15, 10, 5]. The global feature fusion is no need because it only raises a few quality of the image but requires a significant amount of memory and computational cost for 1x1 convolution.

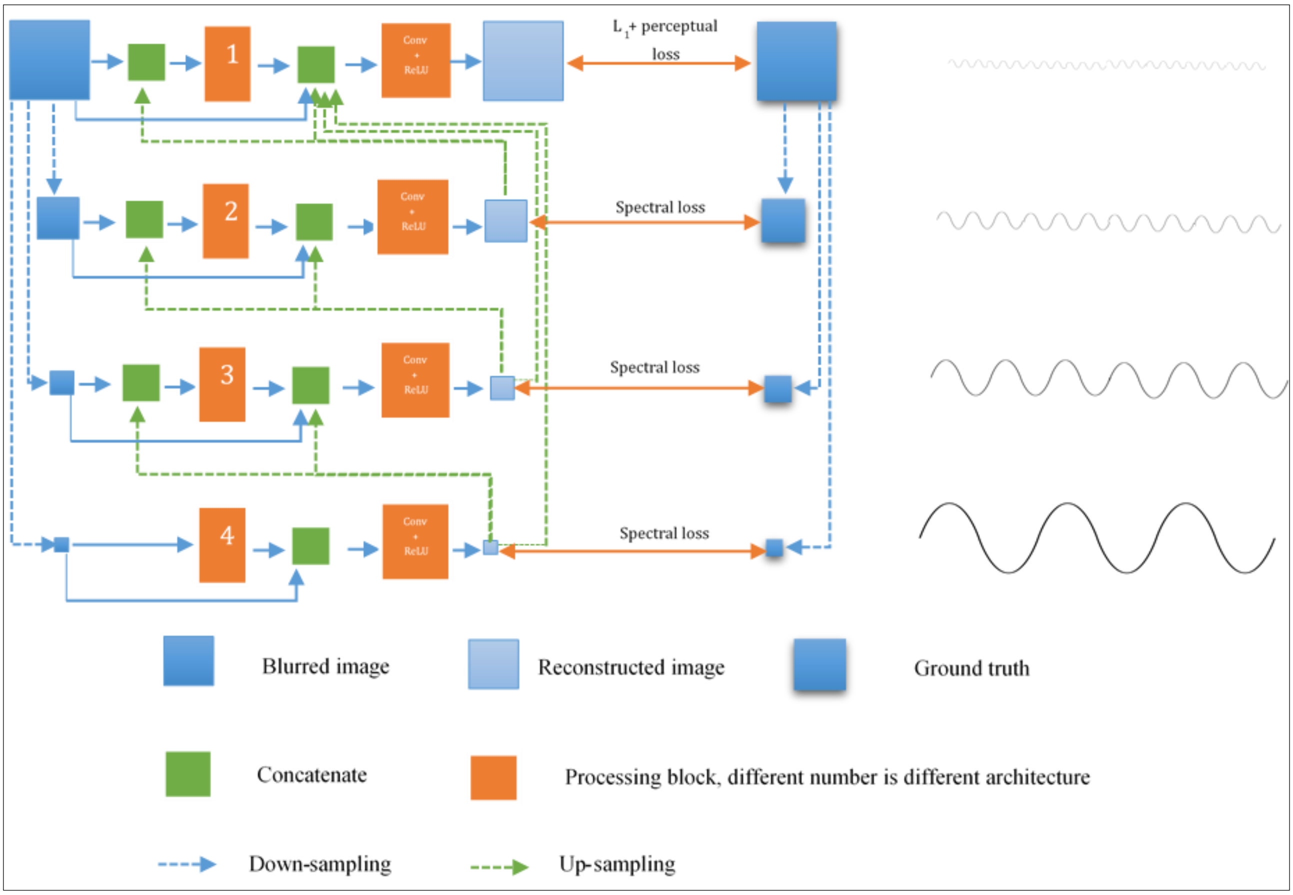

Fig 4: The proposed architecture

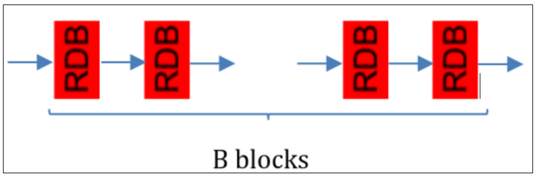

Fig 5: Processing block

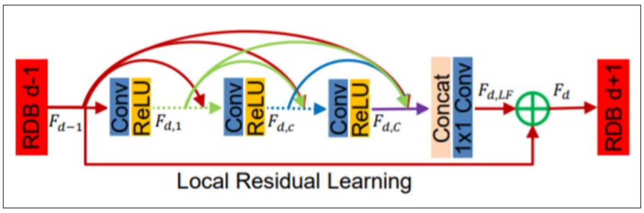

Fig 6: Residual dense block

4. Experiments

4.1 Dataset

I use the dataset captured in [1]. The dataset contains 240 pairs of a clean image and a degraded image. I split it into train/val/test set with a ratio of 0.75/0.1/0.15. For data augmentation, I crop the original image to 256x256 images overlapping 128 pixels. Therefore, from 1024x2048 images, I generate 105 patches having size 256x256. We do not flip, resize the image because the image is the blurring pattern that is fixed for the camera with a fixed orientation.

4.2 Training

The model train from scratch with all weights are randomly initialized. All the scale levels are convergence. It proves that the model utilizes the information from all scales. My training process is stopped at 50 epochs. The trained model is compared with other state-of-the-art architecture in the test set.

4.3 Compare with other SoTa approaches

Our architecture achieves the first place on SSIM and second place on PNSR. It proves that our architecture tends to create a more semantic image as well as keep the pixel loss and low as possible.

Table 1: Compare with other SoTa architectures

Architectures	T-OLED	
	PSNR	SSIM
Unet (basedline)	37.0	0.9313
Unet (basedline) using OctaveConv	36.97	0.9320
DeblurGAN	34.70	0.9114
SRResNet	36.88	0.9360
RDN	**37.21**	0.9375
DeblurGANv2	35.54	0.9084
SRN	37.18	0.9328
Our	37.18	**0.9378**

5. Conclusion

In this paper, we proposed a new approach for restoring the image captured by an under-display camera. The architecture utilizes the multi-scales information for restoring the image. It achieves a good score on SSIM as well and PSNR compared to existed methods.

Acknowledgments

This work is supported by FPT University, Hanoi, Vietnam

Disclosure of conflict of interest

On behalf of all authors, corresponding author declares that there is no conflict of interest to publish this research.

6. References

1. Zhou Y, David R, Emerton N, Lim S, Large T. Image Restoration for Under-Display Camera, Arxiv, 2020.
2. Sun Q, Ethan T, Fu Q, Heidrich W, Heide F. Learning Rank-1 Diffractive Optics for Single-shot High Dynamic Range Imaging, CVPR, 2020.
3. Khan S, Adarsh V, Vivek B, Tan J, Ashok V, Kaushik M. Towards Photorealistic Reconstruction of Highly Multiplexed Lensless Images, ICCV, 2019.
4. Adam K, Fattal R. Deblurring using Analysis-Synthesis Networks pair, 2020.
5. Nah S, Kim TH, Lee KM. Deep Multi-scale Convolutional Neural Network for Dynamic Scene Deblurring, CVPR, 2017.
6. Justin J, Alexandre A, Li F. Perceptual Losses for Real-Time Style Transfer and Super-Resolution. ECCV, 2016.
7. Orest K, Budzan V, Mykola M, Mishkin D, Jiri M. DeblurGAN: Blind motion deblurring using conditional adversarial networks, CVPR, 2018.
8. Orest K, Tetiana M, Wu J, Wang Z. DeblurGAN-v2: Deblurring (Orders-of-Magnitude) Faster and Better, ICCV, 2019.
9. Zhang K, Zuo W, Zhang L. Deep plug-and-play super-resolution for arbitrary blur kernels, CVPR, 2019
10. Lim B, Son SH, Kim HW, Nah S, Lee KM. Enhanced deep residual networks for single image Super-Resolution, CVPR, 2017.
11. Zhang Y, Tian Y, Yu K, Zhong B, Fu Y. Residual dense network for image super-resolution, CVPR 2018.
12. Lim S, Kim J, Kim W. Deep spectral-spatial network for single image deblurring, IEEE Signal Processing Letters, 2020.