UDC 2020 Challenge on Image Restoration of Under-Display Camera: Methods and Results

45 authors, including:

Yuqian Zhou
University of Illinois, Urbana-Champaign
40 PUBLICATIONS **993** CITATIONS

SEE PROFILE

Yang Qirui
Chinese Academy of Sciences
8 PUBLICATIONS **14** CITATIONS

SEE PROFILE

Baopu Li
Baidu Online Network Technology
137 PUBLICATIONS **2,314** CITATIONS

SEE PROFILE

Densen Puthussery
College of Engineering Trivandrum
19 PUBLICATIONS **116** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Project   Photoplethysmography Signal Processing View project

Project   compressive light field reconstruction using deep learning View project

UDC 2020 Challenge on Image Restoration of Under-Display Camera: Methods and Results

Yuqian Zhou, Michael Kwan, Kyle Tolentino, Neil Emerton, Sehoon Lim, Tim Large, Lijiang Fu, Zhihong Pan, Baopu Li, Qirui Yang, Yihao Liu, Jigang Tang, Tao Ku, Shibin Ma, Bingnan Hu, Jiarong Wang, Densen Puthussery, Hrishikesh P S, Melvin Kuriakose, Jiji C V, Varun Sundar, Sumanth Hegde, Divya Kothandaraman, Kaushik Mitra, Akashdeep Jassal, Nisarg A. Shah, Sabari Nathan, Nagat Abdalla Esiad Rahel, Dafan Chen, Shichao Nie, Shuting Yin, Chengconghui Ma, Haoran Wang, Tongtong Zhao, Shanshan Zhao, Joshua Rego, Huaijin Chen, Shuai Li, Zhenhua Hu, Kin Wai Lau, Lai-Man Po, Dahai Yu, Yasar Abbas Ur Rehman,Yiqun Li, Lianping Xing

Abstract. This paper is the report of the first Under-Display Camera (UDC) image restoration challenge in conjunction with the RLQ workshop at ECCV 2020. The challenge is based on a newly-collected database of Under-Display Camera. The challenge tracks correspond to two types of display: a 4k Transparent OLED (T-OLED) and a phone Pentile OLED (P-OLED). Along with about 150 teams registered the challenge, eight and nine teams submitted the results during the testing phase for each track. The results in the paper are state-of-the-art restoration performance of Under-Display Camera Restoration. Datasets and paper are available at https://yzhouas.github.io/projects/UDC/udc.html.

Keywords: Under-Display Camera, Image Restoration, Denoising, Deblurring

1 Introduction

Under-Display Camera (UDC) [34] is specifically designed for full-screen devices as a new product trend, eliminating the need for bezels. Improving the screen-to-body ratio will enhance the interaction between users and devices. Mounting the display in front of a camera imaging lens will cause severe image degradation like low-light and blur. It is then desirable to propose proper image restoration algorithms for a better imaging quality identical to the original lens. It will also potentially benefit the downstream applications like object detection[27] and face analysis[33].

Obtaining such algorithms can be challenging. First, most existing methods leverage the advantages of deep learning to resolve multiple image degradation problems, such as image denoising[2,1,32,31,13], deblurring[8], super-resolution[19,18] etc. Restoration of UDC images, as a problem of recovering combined degradation, requires the joint modeling of methods resolving different optical effects caused by the displays and camera lens. It also requires the

researchers to understand inter-disciplinary knowledge of optics and vision. Second, data acquisition process can be challenging due to variant types of displays and cameras. Collecting data sets consisting of pairs of degraded and undegraded images that are in other respects identical is challenging even using special display-camera combination hardware. Furthermore, the trained model may not be easily generalized to other devices..

In this paper, we report the methods and results from the participants of the first Under-Display Camera Challenge in conjunction with the Real-world Recognition from Low-quality Inputs (RLQ) workshop of ECCV 2020. We held this image restoration challenge to seek an efficient and high-performance image restoration algorithm to be used for recovering under-display camera images. Participants greatly improved the restoration performance compared to the baseline paper. More details will be discussed in the following sections.

2   Under-Display Camera (UDC) Challenge

2.1   UDC Dataset

The UDC dataset is collected using a monitor-based imaging system as illustrated in the baseline paper[34]. Totally 300 images from DIV2K[3] dataset are displayed on the monitor screen, and paired data is recorded using a FLIR Flea camera. In this challenge, we only use the 3-channel RGB data for training and validation. The training data consists of 240 pairs of 1024×2048 images, totally 480 images. Validation and Testing inputs each consist of 30 images of the same resolution. The challenge is organized in two phases: validation and testing. We only release the ground truth of the validation set after the end of the validation phase, while the ground truth of the testing partition is kept hidden.

2.2   Challenge Tracks and Evaluation

The challenge had two tracks: **T-OLED** and **P-OLED** image restoration. Participants were encouraged to submit results on both of them, but only attending one track was also acceptable. For both tracks, we evaluated and ranked the algorithms using the standard Peak Signal To Noise Ratio (PSNR). Additional measurements like Structural Similarity (SSIM) and inference time are also reported for reference. Although an algorithm with high efficiency is extremely important for portable devices, we did not rank the participants based on the inference time. In total of 83 teams took part in the T-OLED track, and 73 teams registered the P-OLED track. Finally, 8 teams submitted the testing results to T-OLED track, and 9 teams to the P-OLED track.

3   Teams and Methods

In this section, we summarize all the methods from the participants who submitted the final results and reports for each track.
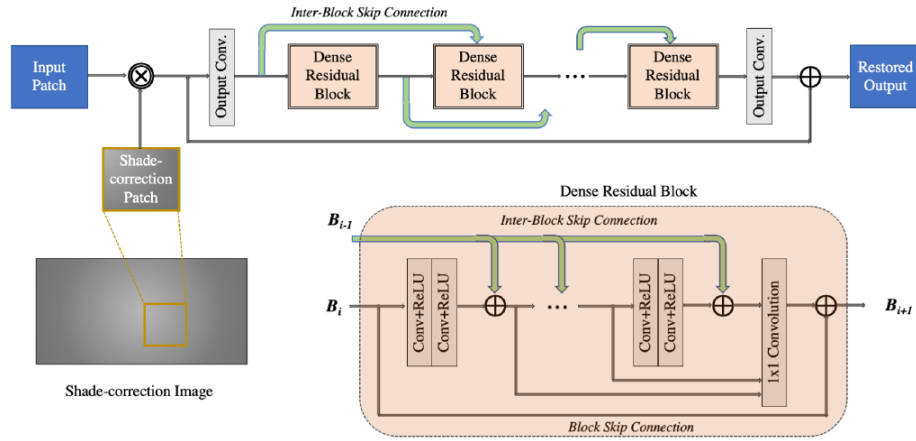
Fig. 1: The Dense Residual Network architecture proposed by Team Baidu Research Vision.

3.1 Baidu Research Vision

Members: Zhihong Pan, Baopu Li
Affiliations: Baidu Research (USA)
Track: T-OLED
Title: Dense Residual Network with Shade-Correction for UDC Image Restoration The architecture proposed by Team Baidu Research Vision is shown in Fig. 1. The team branches off of prior work of a dense residual network for raw image denoising and demosaicking [1] with a newly added novel shade-correction module. The novel shade-correction module consists of a set of learned correction-coefficients with the same dimension as the full-size image. Matching coefficients of the input patch are multiplied with the input for shade correction. The proposed shade-correction module could learn patterns related to the specific T-OLED screen, so additional learning might be needed for different set-ups. However, the team believes this fine-tuning process tends to be efficient. The model is trained on patches of size 128×128.

3.2 BigGuy

Members: Qirui Yang, Yihao Liu, Jigang Tang, Tao Ku
**Affiliations:** Chinese Academy of Sciences, Shenzhen Institutes of Advanced Technology
Track: T-OLED and P-OLED
Title: Residual and Dense U-Net[21] for Under-display Camera Restoration The team's method is based on the U-net network. The U-Net encoder network consists of several residual dense blocks. A decoder network is constructed by skip connection and pixel-shuffle upsampling[22]. Their experiments show
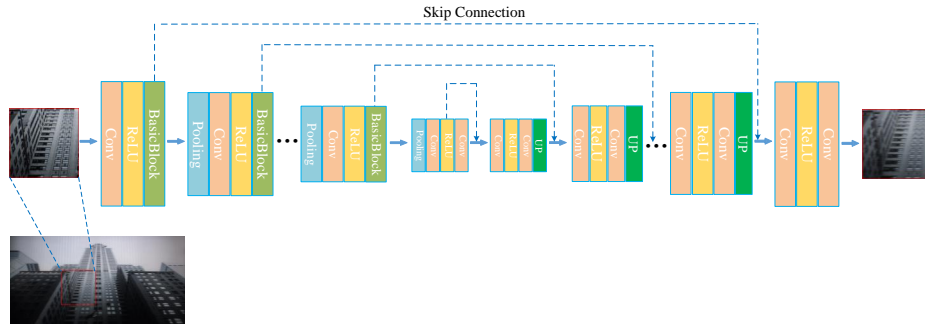
(a) The overall architecture is similar with UNet. We can choose or design different "basic blocks" (e.g. residual block,dense block, residual dense block) to obtain better performance.
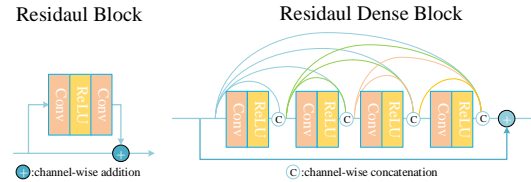
(b) Left: residual block. Right: the proposed residual dense block. It consists of a residual connection and a dense module composed of four convolutional layers.

Fig. 2: The architectures proposed by Team BigGuy.

that T-OLED and P-OLED have different adaptability to the model and the patch size during training. Therefore, they proposed two different U-Net structures. For the T-OLED track, they used Residual Dense Blocks as the basic structure and proposed a Residual-Dense U-Net model (RDU-Net) as shown in Fig. 2.

For the P-OLED track, they found P-OLED panels allow small amounts of light to enter the camera so that P-OLED images present a dull feature. The difference between input and output images of P-OLED dataset is mainly reflected in color and high-frequency texture information. They thus explored residual dense blocks [29], ResNet[9], and RCAB[28] and found residual block achieved the best validation PSNR. The model structure for the P-OLED track is shown in Fig.2.

3.3   BlackSmithM

Members: Shibin Ma, Bingnan Hu, Jiarong Wang
Affiliations: None
Track: P-OLED
Title: P-OLED Image Reconstruction Based on GAN Method In view of the poor quality blurred image of the P-OLED track, the team focused on

adjusting the light first. After the image is dimmed, they can better remove the blur. The team used the pix2pix [12,35] model to adjust the poor image light, and at the same time, it was possible to recover the image information. Before the image passed the pix2pix model, the team preprocessed the single scale Retinex (SSR) [25], and croped the 1024 data set to the left and right images of 1024×1024. The image after the pix2pix network contained noise, so a Gaussian filter was used to process the image to make the resulting image more smooth and real, thus improving the PSNR value of the image.

3.4   CET_CVLab

Members: Densen Puthussery, Hrishikesh P S, Melvin Kuriakose, Jiji C V
Affiliations: College of Engineering, Trivandrum, India
Track: T-OLED and P-OLED
Title: Dual Domain Net (T-OLED) and Wavelet decomposed dilated pyramidal CNN (P-OLED) The team proposed encoder-decoder structures to learn the restoration. For the T-OLED track, they proposed a dual-domain net (DDN) inspired by [30]. In the dual domain method, the image features are processed in both pixel domain and frequency domain using implicit discrete cosine transform. This enables the network to correct the image degradation in both frequency and pixel domain, thereby enhancing the restoration. The DDN architecture is shown in Fig. 3(a).

For the P-OLED track, inspired by the multi-level wavelet-CNN (MWCNN) proposed by Liu et al. [14], they proposed Pyramidal Dilated Convolutional RestoreNet (PDCRN) which follows an encoder-decoder structure as shown in Fig. 3(b). In the proposed network, the downsampling operation in the encoder is discrete wavelet transform (DWT) based decomposition instead of down-sampling convolution or pooling. Similarly, in the decoder network, inverse discrete wavelet transform (IDWT) is used instead of upsampling convolution. In the wavelet based decomposition used here, the information from all the channels are combined in the downsampling process to minimize information loss when compared to that of convolutional downsampling. The feature representation for both tracks is made efficient using a pyramidal dense dilated convolutional block. The dilation rate is gradually decreased as the dilated convolution pyramid is moved up. This is to compensate for information loss that may occur with highly dilated convolution due to a non-overlapping moving window in the convolution process.

3.5   CILab IITM

Members: Varun Sundar, Sumanth Hegde, Divya Kothandaraman, Kaushik Mitra
Affiliations: Indian Institute of Technology Madras
Track: T-OLED and P-OLED
Title: Deep Atrous Guided Filter for Image Restoration in Under Display Cameras. The team uses two-stage pipeline for the task as shown in

(a) The Dual Domain Net for T-OLED Track.

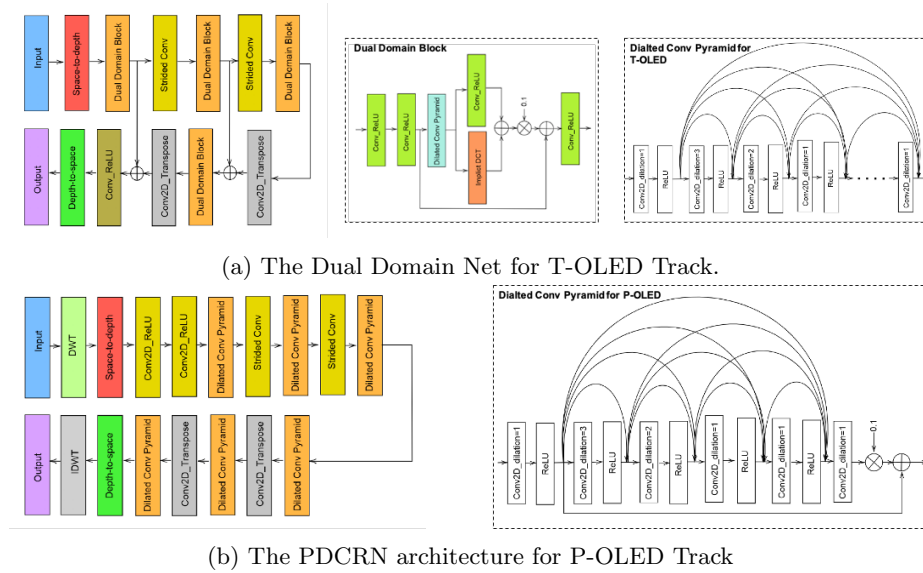(b) The PDCRN architecture for P-OLED Track

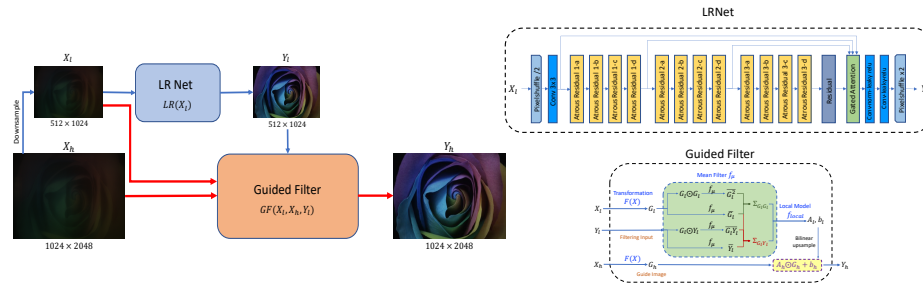Fig. 3: The architectures proposed by Team CET_CVLab.

Fig. 4: The Deep Atrous Guided Filter architectures of the LRNet and the guided filter proposed by Team CILab IITM.

Fig. 4. The first stage is a low-resolution network (LRNet) which restores image quality at low-resolution. The low resolution network retains spatial resolution and emulates multi-scale information fusion with multiple atrous convolution blocks [5,6] stacked in parallel. In the second stage, they leverage a guided filter to produce a high resolution image from the low resolution refined image obtained from stage one. They further propose a simulation scheme to augment data and boost performance. More details are in the team's challenge report[23].

3.6 Hertz

Members: Akashdeep Jassal[1], Nisarg A. Shah[2]
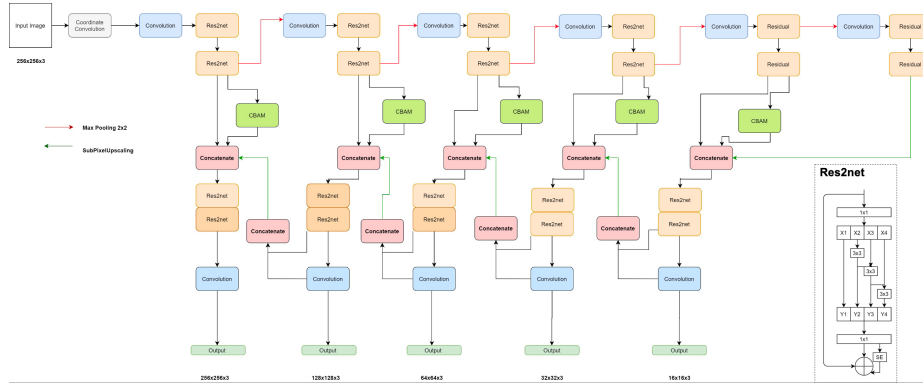**Affiliations:** Punjab Engineering College, India[1], IIT Jodhpur[2]

Fig. 5: The Lightweight Multi Level Supervision Model architecture proposed by Team Image Lab.

Track: P-OLED

Title: P-OLED reconstruction using GridDehazeNet Based on GridDe-hazeNet which aims to clear haze from a low resolution image [16], the team uses the network which contains a pre-processing residual dense block, a grid-like backbone of the same residual dense blocks interconnected with convolutional downsampling and upsampling modules, and a post-processing stage of another residual dense block.

3.7   Image Lab

Members: Sabari Nathan[1], Nagat Abdalla Esiad Rahel[2]
Affiliations: Couger Inc, Japan[1], Al-Zintan University, Libya[2]
Track: T-OLED and P-OLED
Title: Image Restoration using Light weight Multi Level Supervision Model The team proposes a Lightweight Multi Level Supervision Model inspired by [20]. The architecture is shown in Fig. 5. The input image is first passed to the coordinate convolutional layer to map the pixels to a Cartesian coordinate space[15], and then fed into the encoder. The encoder composed of 3×3 convolution layers, two Res2Net[7] blocks, and a downsampling layer, while the decoder block replaces the last component with a subpixel scaling layer[22]. A convolution block attention module (CBAM)[26] in the skip connection is concatenated with the encoding block as well.

3.8   IPIUer

Members: Dafan Chen, Shichao Nie, Shuting Yin, Chengconghui Ma, Haoran Wang
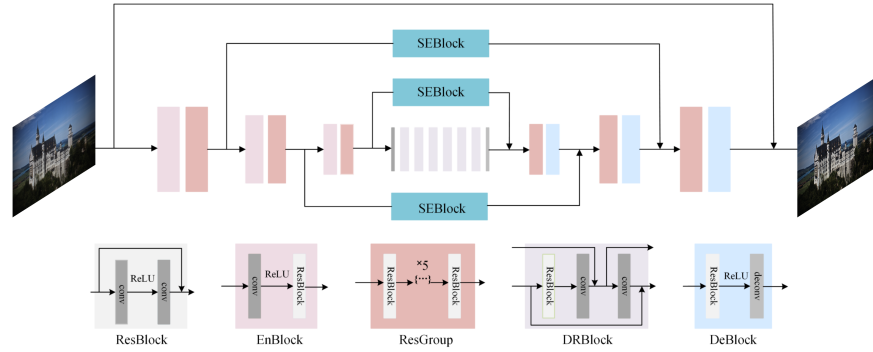**Affiliations:** Xidian University
Track: T-OLED

Fig. 6: The novel Unet structure proposed by Team IPIUer.

Title: Channel Attention Image Restoration Networks with Dual Residual Connection The team proposes a novel UNet model inspired by Dual Residual Networks[17] and Scale-recurrent Networks (SRN-DeblurNet)[24]. As shown in Fig. 6, in the network, there are 3 EnBlocks, 3 DeBlocks and 6 DRBlocks. The entire network has three stages and one bridge between encoder and decoder. Every stage consists of one EnBlock/DeBlock and a residual group (ResGroup). The ResGroup has seven residual blocks (ResBlock). Between the encoder and decoder, a dual residual block (DRBlock) is used to extract high level semantic features effectively.. The skip connection uses squeeze-and-exitation blocks [10] which aims to highlight the features of some dimensions.

3.9   lyl

Members: Tongtong Zhao, Shanshan Zhao
Affiliations: Dalian Maritime University
Track: T-OLED and P-OLED
Title: Coarse to Fine Pyramid Networks for Progressive Image Restoration The team proposes a coarse to fine network (CFN) for progressive reconstruction. Specifically, in each network level, the team proposes a lightweight upsampling module (LUM), also named FineNet as in Fig. 7, to process the input, and merge it with the input features. Such progressive cause-and-effect process helps to achieve the principle for image restoration: high-level information can guide an image to recover a better restored image. The authors claim that they can achieve competitive results with a modest number of parameters.

3.10   San Jose Earthquakes

Members: Joshua Rego, Huaijin Chen, Shuai Li, Zhenhua Hu
Affiliations: SenseBrain Technology
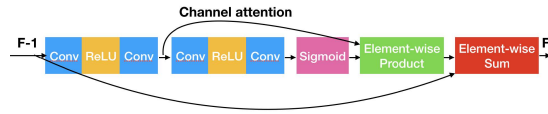**Track**: T-OLED and P-OLED

Fig. 7: The Lightweight Upsampling Module (LUM) named FineNet proposed by Team lyl.
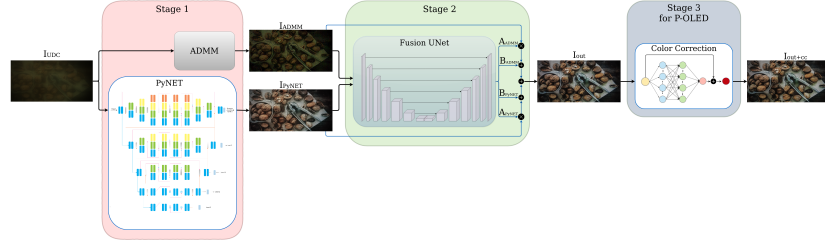

Fig. 8: The network architecture of multi-stage restoration proposed by Team San Jose Earthquakes.

Title: Multi-stage Network for Under-display Camera Image Restoration The team proposes multiple stage networks as shown in Fig. 8 to solve different issues caused by under-display cameras. For T-OLED, the pipeline is two-staged. The first stage uses a multi-scale network PyNET[11] to recover intensity and largely apply deblur to the input image. The second stage is a U-Net fusion network that uses the output of the PyNET as well as a sharper, but noisier, alternating direction method of multipliers (ADMM)[4] reconstruction as the inputs of the network and outputs weights used to combine the two inputs for a sharper, de-noised result. P-OLED pipeline uses an additional third stage color correction network to improve color consistency with the target image by pixel-wise residual learning.

The authors mentioned that training solely through the first-stage network, while greatly restoring towards the target image, was unable to restore sharpness completely, especially in higher frequency textures. The ADMM output, on the other hand, was able to retain high-frequency sharpness, but suffered largely from noise and some additional artifacts. The fusion network blends the desirable characteristics from the two to slightly improve the results. However, one drawback of the method is that ADMM takes about 2.5 mins to solve each image, which is longer for inference. Nevertheless, the method is a novel approach fusing the implementation of a traditional and deep-learning method.

3.11   TCL Research HK

Members: Kin Wai Lau[1,2], Lai-Man Po[1], Dahai Yu[2], Yasar Abbas Ur Rehman[1,2], Yiqun Li[1,2], Lianping Xing[2]
Affiliations: City University of Hong Kong [1], TCL Research Hong Kong [2]
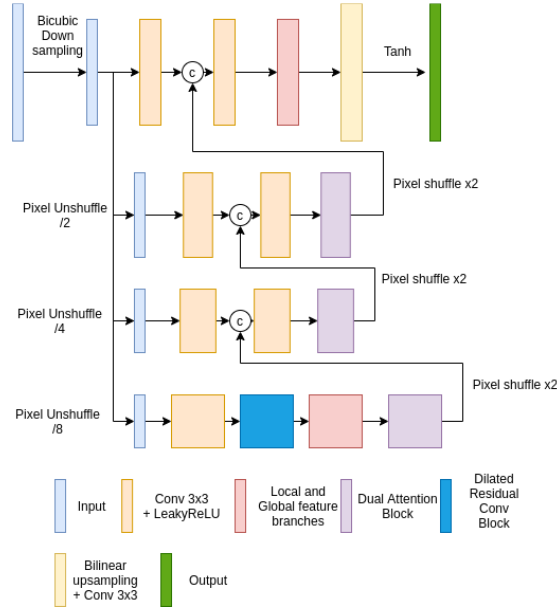
Fig. 9: The network architecture proposed by Team TCL Research HK.

Track: P-OLED

Title: Self-Guided Dual Attention Network for Under Display Camera Image Restoration The team proposes a multi-scale self-guidance neural architecture containing (1) multi-resolution convolutional branch for extracting multi-scale information, (2) low-resolution to high-resolution feature extraction for guiding the intermediate high-resolution feature extraction process, (3) spatial and channel mechanisms for extracting contextual information, (4) Dilated Residual Block (DRB) to increase the receptive field for preserving the details, (5) local and global feature branch for adjusting local information (e.g., contrast, detail, etc.) and global information (e.g., global intensity, scene category, color distribution, etc.) The network architecture proposed by Team TCL Research HK for Self-Guided Dual Attention Network is shown in Fig. 9.

4   Results

This section presents the performance comparisons of all the methods in the above sections. The ranking and other evaluation metrics are summarized in Table 1 for T-OLED track, and in Table 2 for P-OLED.

Top Methods All the submitted methods are mainly deep-learning oriented. On both tracks, top methods achieved very close PSNR. Among the top-3 methods for T-OLED track, directly training a deep model (e.g. modified UNet) in an end-to-end fashion mostly achieved competitive performance. The outperformed

Team	Username	PSNR	SSIM	TT(h)	IT (s/frame)	CPU/GPU	Platform	Ensemble	Loss
Baidu Research Vision	zhihongp	$38.23_{(1)}$	$0.9803_{(1)}$	12	11.8	Tesla M40	PaddlePaddle, PyTorch	flip(×4)	L_1,SSIM
IPIUer	TacoT	$38.18_{(2)}$	$0.9796_{(3)}$	30	1.16	Tesla V100	PyTorch	None	L_1
BigGuy	JaydenYang	$38.13_{(3)}$	$0.9797_{(2)}$	23	0.3548	Tesla V100	PyTorch	-	L_1
CET_CVLAB	hrishikeshps	$37.83_{(4)}$	$0.9783_{(4)}$	72	0.42	Tesla P100	Tensorflow	-	L_2
CILab IITM	ee15b085	$36.91_{(5)}$	$0.9734_{(6)}$	96	1.72	GTX 1080 Ti	PyTorch	flip(×4) model(×8)	L_1
lyl	tongtong	$36.72_{(6)}$	$0.9776_{(5)}$	72	3	-	PyTorch	model(-)	L_1
Image Lab	sabarinathan	$34.35_{(7)}$	$0.9645_{(7)}$	-	1.6	GTX 1080 Ti	Keras	model(-)	L_2,SSIM
San Jose Earthquakes	jdrego	$33.78_{(8)}$	$0.9324_{(8)}$	18	180	-	PyTorch	model(-)	-

Table 1: Results and rankings of methods of T-OLED Track. TT: Training Time. IT: Inference Time

Team	Username	PSNR	SSIM	TT(h)	IT (s/frame)	CPU/GPU	Platform	Ensemble	Loss
CET_CVLAB	Densen	$32.99_{(1)}$	$0.9578_{(1)}$	72	0.044	Tesla T4	Tensorflow	-	L_2
CILab IITM	varun19299	$32.29_{(2)}$	$0.9509_{(2)}$	96	1.72	GTX 1080 Ti	PyTorch	flip(×4) model(×8)	L_1
BigGuy	JaydenYang	$31.39_{(3)}$	$0.9506_{(3)}$	24	0.2679	Tesla V100	PyTorch	model	L_1
TCL Research HK	stevenlau	$30.89_{(4)}$	$0.9469_{(5)}$	48	1.5	Titan Xp	PyTorch	-	L_1,SSIM
BlackSmithM	BlackSmithM	$29.38_{(5)}$	$0.9249_{(6)}$	-	2.43	Tesla V100	PyTorch	-	L_1
San Jose Earthquakes	jdrego	$28.61_{(6)}$	$0.9489_{(4)}$	18	180	-	PyTorch	model	-
Image Lab	sabarinathan	$26.60_{(7)}$	$0.9161_{(7)}$	-	1.59	-	-	None	L_2,SSIM
Hertz	akashdeepjassal	$25.72_{(8)}$	$0.9027_{(8)}$	-	2.29	Tesla K80	PyTorch	None	VGG
lyl	tongtong	$25.46_{(9)}$	$0.9015_{(9)}$	72	3	-	PyTorch	model(-)	L_1

Table 2: Results and rankings of methods of P-OLED Track. TT: Training Time. IT: Inference Time

results further demonstrate the effectiveness of using UNet embedding Residual Blocks shared by the teams. Similar structures are widely used in image denoising or deblurring. T-OLED degraded images contain blur and noisy patterns due to diffraction effects, which could be the reason for the superiority of directly applying deblurring/denoising approaches. For the P-OLED track, the winner team, CET_CVLab, proposed to use discrete wavelet transform (DWT) to replace the upsampling and downsampling modules. The CILab IITM team proposed a two-stage pipeline with differentiable guided filter for training. The performance gain can also come from the model pre-trained on the simulated data by using measurements provided by the baseline paper[34]. The BigGuy team conducted an extensive model search to find the optimal structures for P-OLED images. Some methods proposed by other teams, though not ranked on top-3, are also novel and worthy to mention. The team San Jose Earthquakes proposed to combine the results of the deep-learning and traditional methods by leveraging the benefits from both ends. The multi-level supervision model proposed by the Image Lab team restores the image in a progressive way. Similarly, the lyl team also share the progressive idea.

In addition to module design, the model depth, parameter amounts, data augmentation and normalization, or training strategies can also cause performance differences. Most teams also report the performance gains from model or inputs ensemble strategies.

T-OLED v.s. P-OLED According to the experiment results, T-OLED demonstrates an easier task than P-OLED. The T-OLED restoration problem itself resembles a combination of deblurring and denoising tasks. However, imag-

ing through P-OLED suffers heavily from lower light transmission and color shift. Some teams like CILab IITM, Image Lab and lyl, which participated in both tracks, chose to use the same models for two tracks, while other teams tend to use different model structures. Team BigGuy explored different module options to better resolve the low-light and color issues of P-OLED inputs. Team CET_CVLab addresses the information loss issues of downsampling from P-OLED inputs by using a wavelet-based decomposition. Team San Jose Earthquakes added an additional color correction stage for P-OLED.

Inference Efficiency We did not rank the methods by the inference time due to device and platform difference of different methods. Most methods run about 1 to 3 seconds per image of size 1024×2048 on GPUs. Without further optimization, these models may not be easily applied in mobile devices or laptops for real-time inference of live streams or videos. However, it is still feasible to restore degraded images or videos in an offline way. Team San Jose Earthquakes run a longer inference time since their method involves an additional ADMM optimization process. Team CET_CVLab claimed to achieve 0.044s inference time on a better GPU, which makes the method both outperformed and high-efficient.

5   Conclusions

We summarized and presented the methods and results of the first image restoration challenge on Under-Display Camera (UDC). The testing results represent state-of-the-art performance on Under-Display Camera Imaging. Participants extensively explored state-of-the-art deep-learning based architectures conventionally used for image restoration. Some additional designs like shade and color correction are also proven beneficial to be adaptive to the through-display imaging tasks. However, the results are specifically limited to two display types, P-OLED and T-OLED, and a single camera. This further suggests the need for exploring more display types and hardware set-ups so the model can be generalized.

References

1. Abdelhamed, A., Afifi, M., Timofte, R., Brown, M.S.: Ntire 2020 challenge on real image denoising: Dataset, methods and results. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. pp. 496–497 (2020)
2. Abdelhamed, A., Timofte, R., Brown, M.S.: Ntire 2019 challenge on real image denoising: Methods and results. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. pp. 0–0 (2019)
3. Agustsson, E., Timofte, R.: Ntire 2017 challenge on single image super-resolution: Dataset and study. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. pp. 126–135 (2017)
4. Boyd, S., Parikh, N., Chu, E.: Distributed optimization and statistical learning via the alternating direction method of multipliers. Now Publishers Inc (2011)
5. Brehm, S., Scherer, S., Lienhart, R.: High-resolution dual-stage multi-level feature aggregation for single image and video deblurring. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. pp. 458–459 (2020)
6. Chen, D., He, M., Fan, Q., Liao, J., Zhang, L., Hou, D., Yuan, L., Hua, G.: Gated context aggregation network for image dehazing and deraining. In: 2019 IEEE Winter Conference on Applications of Computer Vision (WACV). pp. 1375–1383. IEEE (2019)
7. Gao, S., Cheng, M.M., Zhao, K., Zhang, X.Y., Yang, M.H., Torr, P.H.: Res2net: A new multi-scale backbone architecture. IEEE transactions on pattern analysis and machine intelligence (2019)
8. Guo, Q., Feng, W., Chen, Z., Gao, R., Wan, L., Wang, S.: Effects of blur and deblurring to visual object tracking. arXiv preprint arXiv:1908.07904 (2019)
9. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 770–778 (2016)
10. Hu, J., Shen, L., Sun, G.: Squeeze-and-excitation networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 7132–7141 (2018)
11. Ignatov, A., Van Gool, L., Timofte, R.: Replacing mobile camera isp with a single deep learning model. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. pp. 536–537 (2020)
12. Isola, P., Zhu, J.Y., Zhou, T., Efros, A.A.: Image-to-image translation with conditional adversarial networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 1125–1134 (2017)
13. Liu, J., Wu, C.H., Wang, Y., Xu, Q., Zhou, Y., Huang, H., Wang, C., Cai, S., Ding, Y., Fan, H., et al.: Learning raw image denoising with bayer pattern unification and bayer preserving augmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. pp. 0–0 (2019)
14. Liu, P., Zhang, H., Zhang, K., Lin, L., Zuo, W.: Multi-level wavelet-cnn for image restoration. In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops. pp. 773–782 (2018)
15. Liu, R., Lehman, J., Molino, P., Such, F.P., Frank, E., Sergeev, A., Yosinski, J.: An intriguing failing of convolutional neural networks and the coordconv solution. In: Advances in Neural Information Processing Systems. pp. 9605–9616 (2018)
16. Liu, X., Ma, Y., Shi, Z., Chen, J.: Griddehazenet: Attention-based multi-scale network for image dehazing. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 7314–7323 (2019)

17. Liu, X., Suganuma, M., Sun, Z., Okatani, T.: Dual residual networks leveraging the potential of paired operations for image restoration. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 7007–7016 (2019)
18. Mei, Y., Fan, Y., Zhang, Y., Yu, J., Zhou, Y., Liu, D., Fu, Y., Huang, T.S., Shi, H.: Pyramid attention networks for image restoration. arXiv preprint arXiv:2004.13824 (2020)
19. Mei, Y., Fan, Y., Zhou, Y., Huang, L., Huang, T.S., Shi, H.: Image super-resolution with cross-scale non-local attention and exhaustive self-exemplars mining. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 5690–5699 (2020)
20. Nathan, D.S., Beham, M.P., Roomi, S.: Moire image restoration using multi level hyper vision net. arXiv preprint arXiv:2004.08541 (2020)
21. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention. pp. 234–241. Springer (2015)
22. Shi, W., Caballero, J., Huszár, F., Totz, J., Aitken, A.P., Bishop, R., Rueckert, D., Wang, Z.: Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 1874–1883 (2016)
23. Sundar, V., Hegde, S., Kothandaraman, D., Mitra, K.: Deep atrous guided filter for image restoration in under display cameras (2020)
24. Tao, X., Gao, H., Shen, X., Wang, J., Jia, J.: Scale-recurrent network for deep image deblurring. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 8174–8182 (2018)
25. Wei, C., Wang, W., Yang, W., Liu, J.: Deep retinex decomposition for low-light enhancement. arXiv preprint arXiv:1808.04560 (2018)
26. Woo, S., Park, J., Lee, J.Y., So Kweon, I.: Cbam: Convolutional block attention module. In: Proceedings of the European conference on computer vision (ECCV). pp. 3–19 (2018)
27. Yu, X., Gong, Y., Jiang, N., Ye, Q., Han, Z.: Scale match for tiny person detection. In: The IEEE Winter Conference on Applications of Computer Vision. pp. 1257–1265 (2020)
28. Zhang, Y., Li, K., Li, K., Wang, L., Zhong, B., Fu, Y.: Image super-resolution using very deep residual channel attention networks. In: Proceedings of the European Conference on Computer Vision (ECCV). pp. 286–301 (2018)
29. Zhang, Y., Tian, Y., Kong, Y., Zhong, B., Fu, Y.: Residual dense network for image restoration. IEEE Transactions on Pattern Analysis and Machine Intelligence (2020)
30. Zheng, B., Chen, Y., Tian, X., Zhou, F., Liu, X.: Implicit dual-domain convolutional network for robust color image compression artifact reduction. IEEE Transactions on Circuits and Systems for Video Technology (2019)
31. Zhou, Y., Jiao, J., Huang, H., Wang, J., Huang, T.: Adaptation strategies for applying awgn-based denoiser to realistic noise. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 33, pp. 10085–10086 (2019)
32. Zhou, Y., Jiao, J., Huang, H., Wang, Y., Wang, J., Shi, H., Huang, T.: When awgn-based denoiser meets real noises. arXiv preprint arXiv:1904.03485 (2019)
33. Zhou, Y., Liu, D., Huang, T.: Survey of face detection on low-quality images. In: 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018). pp. 769–773. IEEE (2018)
34. Zhou, Y., Ren, D., Emerton, N., Lim, S., Large, T.: Image restoration for under-display camera. arXiv preprint arXiv:2003.04857 (2020)

35. Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proceedings of the IEEE international conference on computer vision. pp. 2223–2232 (2017)