A Well-aligned Dataset for Learning Image Signal Processing on Smartphones from a High-end Camera

Yazhou Xing, Changlin Li, Xuaner Zhang, Qifeng Chen

香港科技大學
THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

SIGGRAPH 2022
VANCOUVER+ 8-11 AUG

We collect a large-scale sub-pixel aligned dataset with raw and RGB data pairs captured by two popular smartphones and one high-end camera. Our dataset can be used for learning ISPs to replace the sub-optimal built-in ISPs of smartphones.

PROBLEM

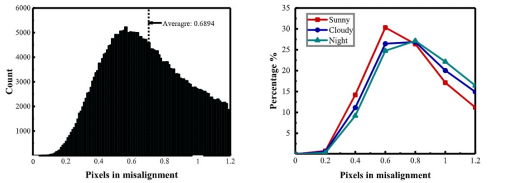
- Not every camera is equipped with an excellent image signal processing (ISP) pipeline that converts raw sensor data into color images.
- It is labor-intensive and challenging to design an ISP pipeline with many independent modules. and thus the ISP on most smartphones is sub-optimal, even for the highly-rated ones such as iPhone.

DATA COLLECTION

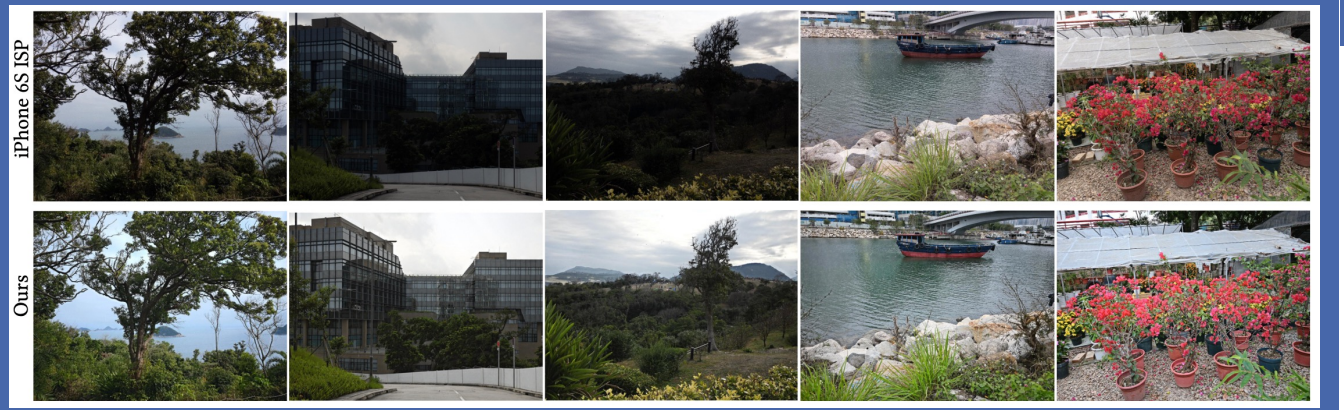
- Over 2000 pairs for training and testing.

iPhone 6S | Nikon Z6 | Camera Rig

DATA ANALYSIS



(a) Misalignment distribution (b) Dependence on illumination



APPROACH

Given an n-pixel input image $X = (X_1, X_2, \dots, X_n), X_i \in R^c$, the output $\hat{Y} = (\hat{Y}_1, \hat{Y}_2, \dots, \hat{Y}_n), \hat{Y}_i \in R^{c'}$ can be obtained by convolution operation with per-pixel kernel W:

$$\hat{Y}_i = \sum_{j \in \delta_i} W_i[i-j] X_j + b,$$

where W^i denotes the conditional convolution kernel at position i, δ_i is the neighborhood window centered at position i. In general, W^i should be a function of input content:

$$W\hat{}i = f(X, \delta_i).$$

We leverage the approximation power of neural networks to estimate the function f.

RESULTS

We provide the quantitative results with baselines. Overall, all perceptual metrics show that our proposed ISP model outperforms the baselines.

	Mi 3			iPhone 6S		
Method	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑
Built-in ISP	0.261	18.82	0.632	0.262	19.65	0.626
DPED	0.474	18.71	0.665	0.453	18.46	0.667
PAC	**0.182**	20.18	0.699	0.204	18.50	0.650
SID	0.264	20.56	0.690	0.295	**21.23**	0.731
Ours	**0.182**	**21.22**	**0.725**	**0.134**	21.09	**0.733**

REFERENCES

1. Chen Chen, Qifeng Chen, Jia Xu, and Vladlen Koltun. 2018. Learning to See in the Dark. In Conference on Computer Vision and Pattern Recognition (CVPR).
2. Hang Su, Varun Jampani, Deqing Sun, Orazio Gallo, Erik Learned-Miller, and Jan Kautz. 2019. Pixel-Adaptive Convolutional Neural Networks. In Conference on Computer Vision and Pattern Recognition (CVPR).
3. Andrey Ignatov, Nikolay Kobyshev, Radu Timofte, Kenneth Vanhoey, and Luc Van Gool. 2017. DSLR-Quality Photos on Mobile Devices with Deep Convolutional Networks. In International Conference on Computer Vision (ICCV).

Download the Dataset

Yazhou Xing:
https://yzxing87.github.io/
yxingag@connect.ust.hk
Qifeng Chen
https://cqf.io/
cqf@ust.hk