HighEr-Resolution Network for Image Demosaicing and Enhancing

Kangfu Mei[1], Juncheng Li[2], Jiajie Zhang[3], Haoyu Wu[1], Jie Li[1], Rui Huang[14]

[1] The Chinese University of Hong Kong, Shenzhen , [2] East China Normal University

[3]Kuaishou Technology, [4]Shenzhen Institute of Artificial Intelligence and Robotics for Society

香港中文大學(深圳)
The Chinese University of Hong Kong, Shenzhen

ICCV 2019 Seoul, Korea

Problem and Contribution

Goal: Learn to restore RGB images generated by DSLR from unprocessed-RAW images generated by smartphones (AIM2019 RAW to RGB Challenge).

Huawei P20 RAW | Huawei P20 RAW (Visualized) | Canon 5D Mark IV

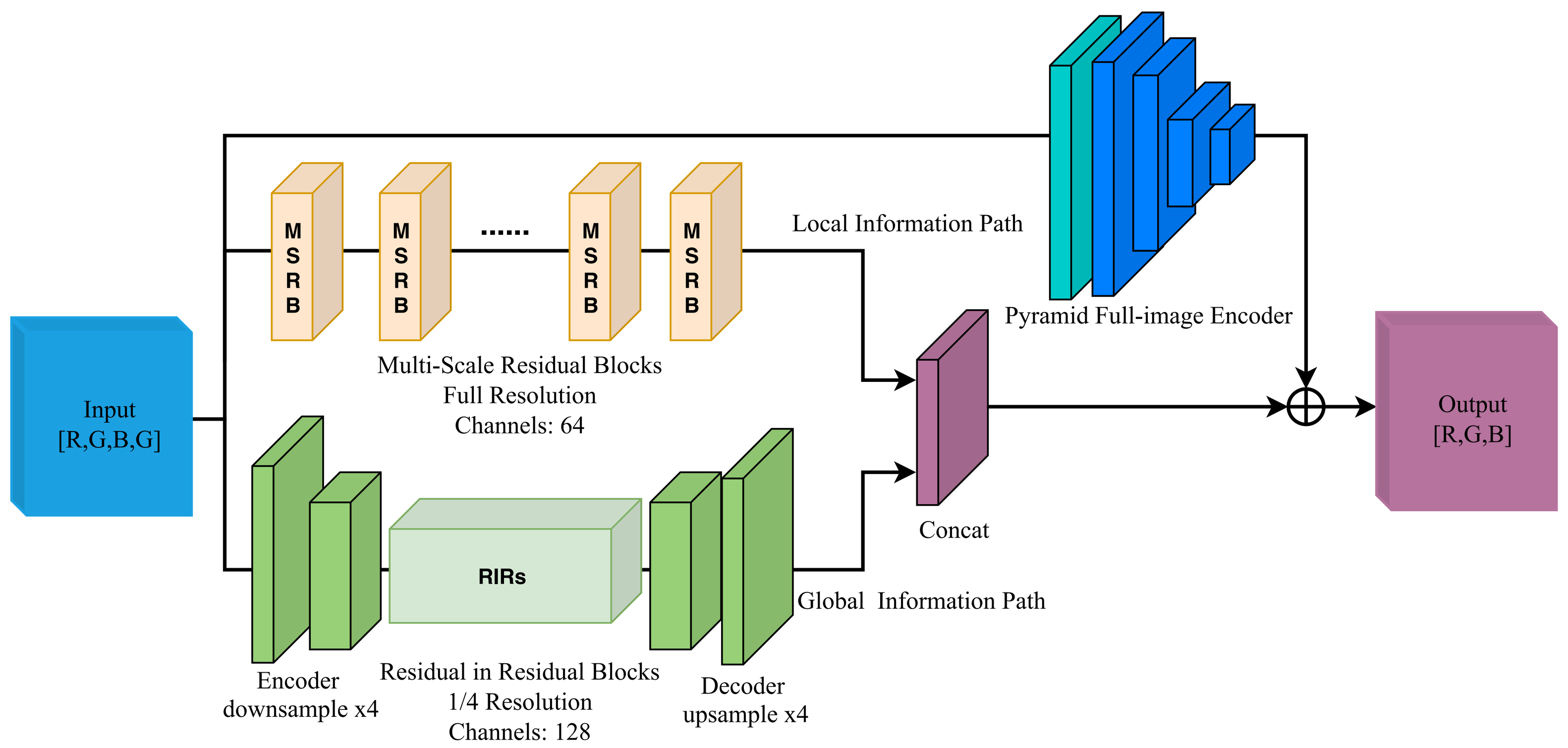
Key Contributions:

- We propose a HighEr-Resolution Network (HERN), which can fully learning local and global information in high-resolution image patches.
- We propose a progressive training method to solve the instability issue and accelerate model convergence.
- Our HERN won second place on track 1 (Fidelity) and won first place on track 2 (Perceptual) in the AIM2019 RAW to RGB Mapping Challenge.

Method

Network Architecture: The proposed HERN is a dual-path network, which consists of a global information path and a local information path.

- The global Information Path in green is based on RCAN but without attention unit. Besides, the autoencoder mechanism is used before the input features and after output features to enlarge the receptive field as well as reduce computational complexity.
- The Local Information Path in yellow is based on MSRN. It aims to extract local information such as textures and edges that destroyed in the encoder downsampling of the global information path.
- Further, a Pyramid Full-Image Encoder that extracts high-level characteristics in fix-resolution (192*192 px) is used to regularize local artifacts, which may be caused by resolution un-consistent.



Loss function:

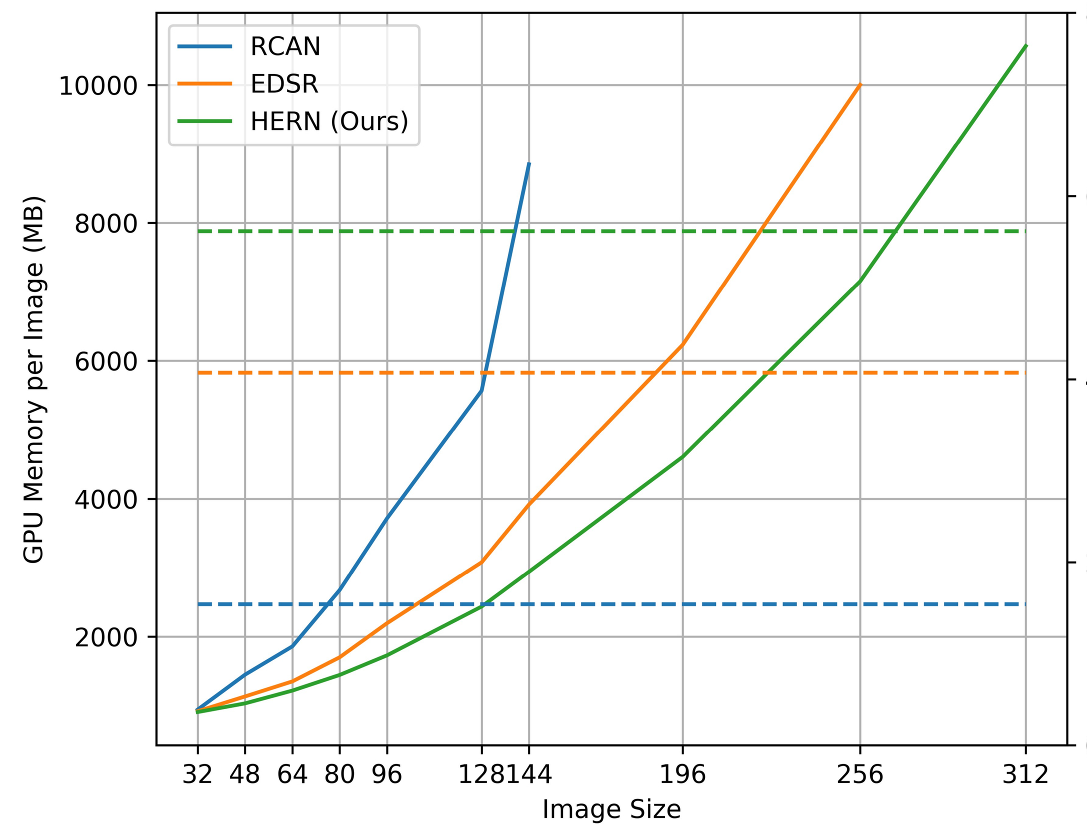
$$\mathcal{L}_1(\theta) = \frac{1}{N}\sum_{n=1}^{N} |\Phi(\mathrm{x_i};\theta) - \mathrm{y_i}|\,, \tag{1}$$

where Φ is the network function and θ represents the learnable parameters, N is the batch size, and $\mathrm{x_i}$, $\mathrm{y_i}$ are the patch pairs of RAW image and RGB image.

Experiments & Results

Provided Datasets for Training: The provided datasets ZRR for traning contains 89000 train image patches of size 224x224 pixels. Some samples is visualized at below:

GPU Memory Usage in Differen Resolution: The GPU memory usage will increases when the resolution of image patches increase. Therefore, it is unrealistic to train state-of-the-art methods on one GPU card with image patches of 224 * 224 pixels



AIM2019 RAW to RGB Mapping Challenge Results

Track	Rank	Team	Metrics	Performance
Track 1	1	1^{st}	PSNR / SSIM	**22.59 / 0.81**
	2	**HERN**	PSNR / SSIM	*22.24 / 0.80*
	3	3^{st}	PSNR / SSIM	21.94 / 0.79
	4	4^{st}	PSNR / SSIM	21.91 / 0.79
	5	5^{st}	PSNR / SSIM	20.85 / 0.77
Track 2	1	**HERN**	MOS	**1.24**
	2	2^{st}	MOS	*1.28*
	3	3^{st}	MOS	1.46
	4	4^{st}	MOS	1.56
	5	5^{st}	MOS	1.92

Artifacts by Resolution Un-Consistent: (a) shows the local artifacts in the wall area, which is generated by the model without the Pyramid Full-Image Encoder. (b) shows clear results in the wall area, which is generated by the model with the Pyramid Full-Image Encoder. The used image is full-resolution (2052*1509 px) while the network is trained in resolution of 224*224 px.

(a) artifacts otuput | (b) clear output

Visual Comparison of Reconstructed Images Trained Using Different Input Patch Sizes

Abalation Study on Using Different Ensemble Strategy During Testing

Epoch	Self-Ensemble	Epoch-Ensemble	PSNR (dB) / SSIM
#113			23.105 / 0.810
#113	✓		23.201 / 0.818
#115			23.162 / 0.811
#115	✓		23.254 / 0.818
#113 & #115	✓	✓	23.304 / 0.818

Scan the Qr-Code:
Github & Preprint & Contact Us