Inverting Image Signal Processing Pipeline with Diffusion Models

Xinman Liu, Xuanchi Ren, Ziyi Wu
University of Toronto

Motivation

- **Raw images** are valuable not only for various image editing tasks but also for computer vision tasks such as image denoising. However, they are memory-intensive.
- **Diffusion models** are good at conditional generation tasks such as image super-resolution and image inpainting.
- Take advantage of the incredible generative power of diffusion models to solve **raw image reconstruction task** as a general **image-to-image translation task**.

Raw Image Reconstruction

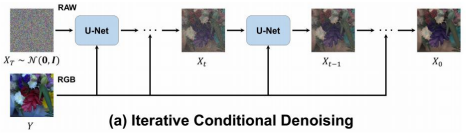
New Technique

Contributions:

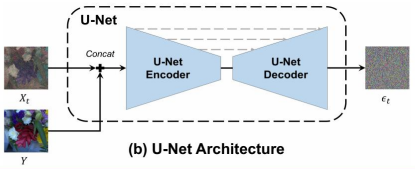
- The first attempt in RAW image reconstruction via diffusion models.
- Study the quality-speed trade-off with different sampling algorithms because diffusion models are slow and memory-consuming in the test-time generation process.

Methodology:

- (a) shows the iterative process RAW image reconstruction procedure, where a U-net performs denoising conditioned on the RGB image.
- (b) illustrates the architecture of U-net with skip-connections, which takes in the concatenation of a noisy RAW image X_t and an RGB image Y, and predict the noise ε_t, which is then used to obtain X_{t-1}.



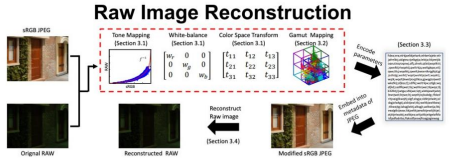
(a) Iterative Conditional Denoising



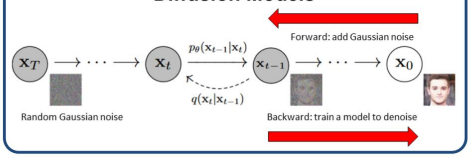
(b) U-Net Architecture

Related Work

Raw Image Reconstruction



Diffusion Models



References

[1] R. M. Nguyen and M. S. Brown, "Raw image reconstruction using a self-contained srgb–jpeg image with small memory overhead," IJCV, 2018.
[2] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," in NeurIPS, 2020.
[3] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in MICCAI, 2015.
[4] C. Saharia, W. Chan, H. Chang, C. Lee, J. Ho, T. Salimans, D. Fleet, and M. Norouzi, "Palette: Image-to-image diffusion models," in ACM SIGGRAPH 2022 Conference Proceedings, 2022.
[5] J. Ho, T. Salimans, A. Gritsenko, W. Chan, M. Norouzi, and D. J. Fleet, "Video diffusion models," in NeurIPS, 2022.
[6] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer, "High-resolution image synthesis with latent diffusion models," in CVPR, 2022.

Experimental Results

Fig. 4. Visualization of our result on NIKON D700 dataset. Our diffusion-based method is capable of synthesizing plausible RAW images.

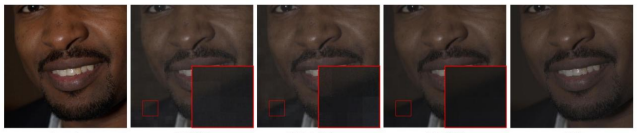


Fig. 5. Qualitative comparison between different sampling strategies. Though DDPM consumes much more time than other sampling strategies, it achieves the best RAW reconstruction quality.

TABLE 1
Quantitative evaluation between our method and baselines.

Method	PSNR↑	
	NIKON D700	Canon EOS 5D
UPI	30.12	26.31
CycleISP	30.19	34.48
InvGrayscale	33.28	38.00
U-Net	41.17	41.14
InvISP	**44.19**	**45.73**
Ours	40.10	41.41

TABLE 2
Ablation study on model design on NIKON D700 subset.

Method	PSNR ↑
Ours (Full Model)	**40.10**
Residual block number $N_{res} = 2$	40.00
U-Net base channel $N_c = 32$	12.84
DDPM variance schedule $\{\beta_t\}_{t=1}^T$	37.61
Patch size 128×128	15.41

TABLE 3
Ablation study on sampling strategy of diffusion models. All speed is measured on 8 NVIDIA V100 GPUs with a batch size of 512 patches.

Method	PSNR ↑	Time per Image (s) ↓
Ours (DDPM)	**40.10**	421.76
DDIM	27.71	166.12
DPM-Solver	28.49	**10.34**