Learning to Generate Realistic Noisy Images via Pixel-level Noise-aware Adversarial Training

NeurIPS 2021

Yuanhao Cai[1,2], Xiaowan Hu[1,2], Haoqian Wang[1,2,*]
Yulun Zhang[3], Hanspeter Pfister[4], and Donglai Wei[5]

Shenzhen International Graduate School, Tsinghua University[1]
The Shenzhen Institute of Future Media Technology[2]
ETH Zürich[3], Harvard University[4], Boston College[5]

Overview

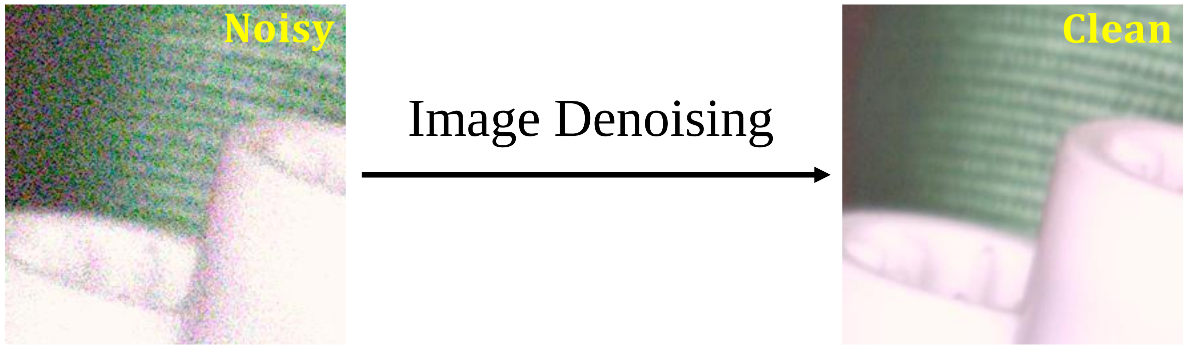
- Introduction
- Method
- Experiments

Introduction

Introduction

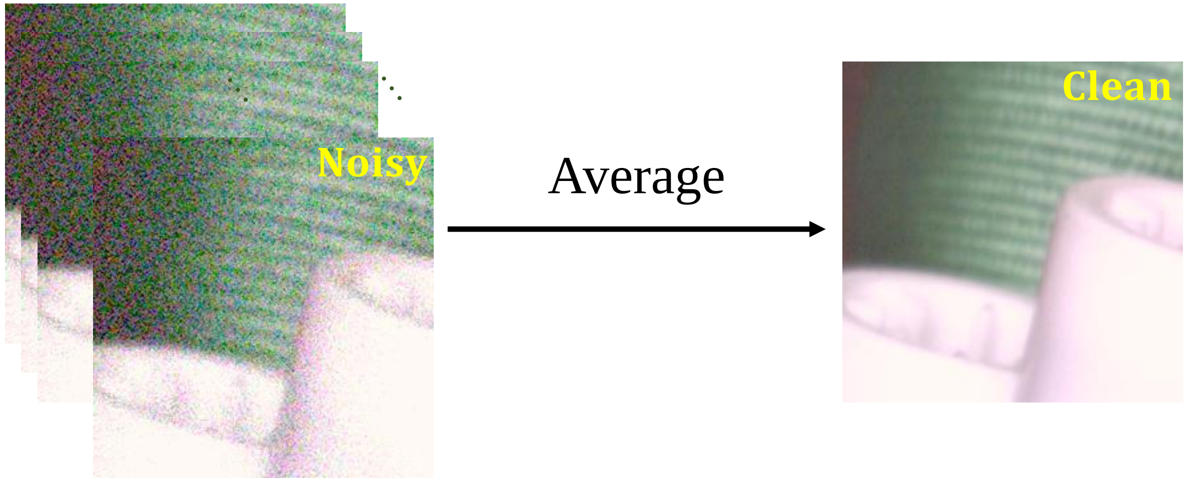
- Image Denoising
  - Traditional methods

    Based on hand-crafted assumptions or models, poor representing capacity



  - Deep learning methods

    Powerful learning models – CNN. Collecting image pairs is tedious and labor-intensive. DL methods face a severe data-hungry situation.



Introduction

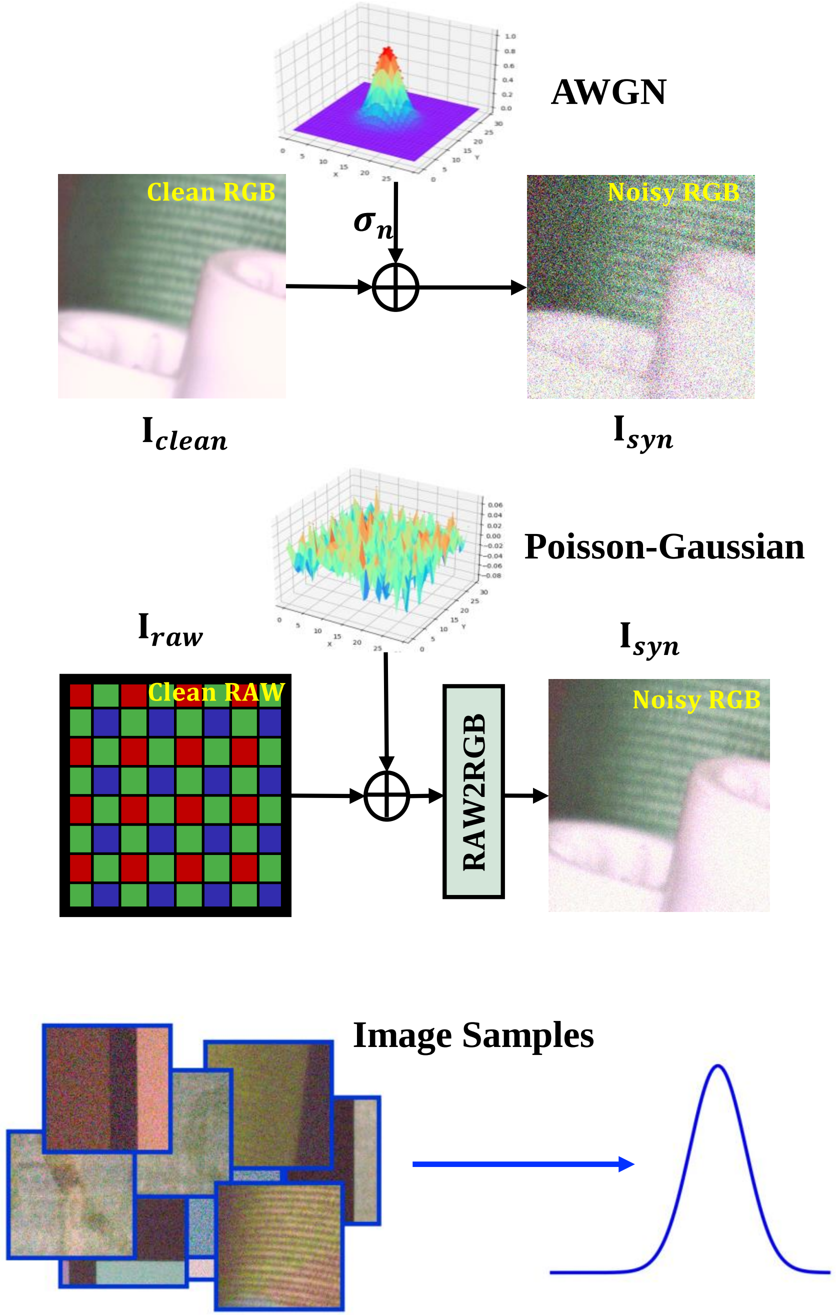
- Noise Generation
  - Gaussian Distribution

    Fundamentally different from real noise. Dramatic performance drop.
  - Pseudo ISP

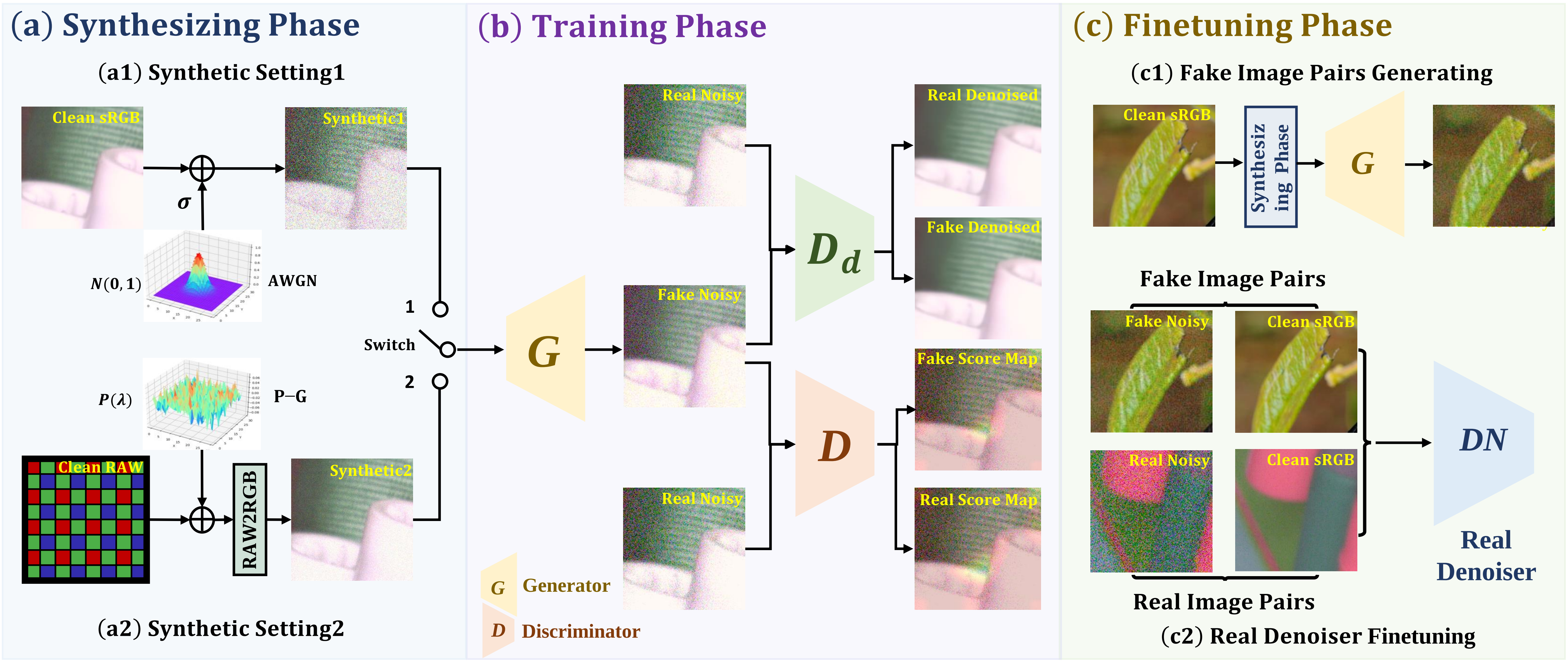
    Cannot ensure the mapping from Poisson-Gaussian noisy RAW image to real-camera noisy RGB image.
  - GAN-based

    Treat images as samples, coarse learning

    There still remains a domain discrepancy



Method



Setting1 : Gaussian Distribution

Setting2 : Poisson-Gaussian Distribution + Pseudo ISP

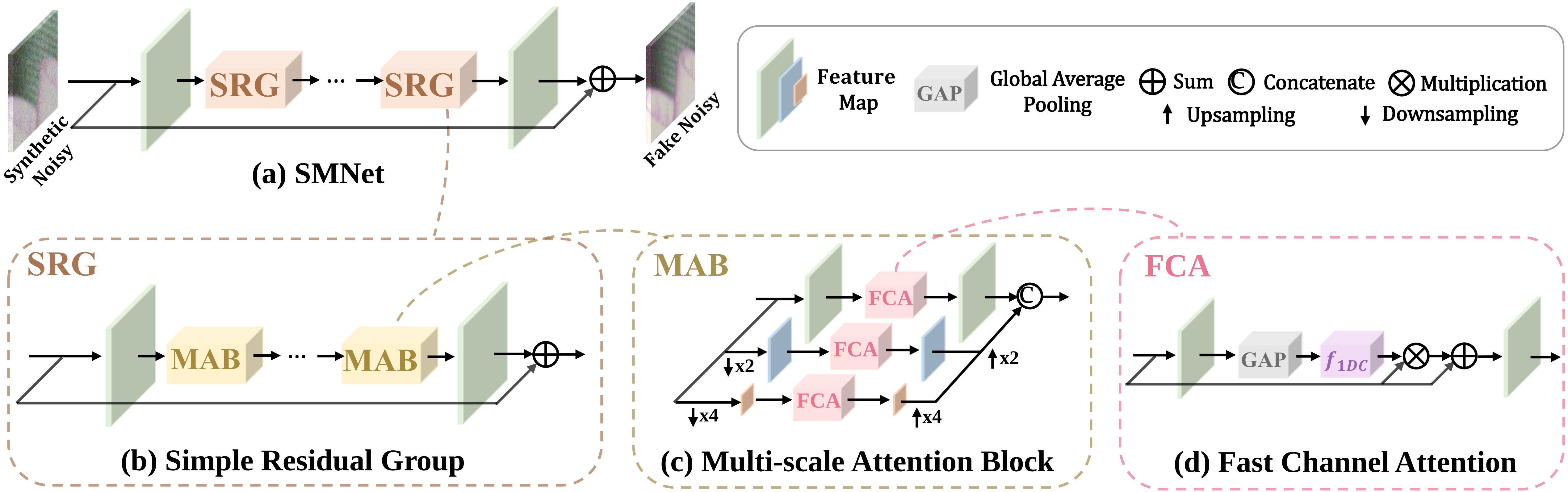
$\boldsymbol{G}$: Noise Generator

$\boldsymbol{D}$: Pixel-level Discriminator

$\boldsymbol{D_d}$: Pre-trained Denoiser

Use $\boldsymbol{G}$ to generate image pairs

Add image pairs to the data pool of any real denoiser $\boldsymbol{DN}$



The Architecture of the Generator

- Simple Multi-scale Network (SMNet) is cascaded by Simple Residual Groups (SRG)
- SRG is built by Multi-scale Attention Blocks (MAB)
- MAB exploits Fast Channel Attention (FCA) mechanism
- FCA uses 1D convolution to cut down Params and FLOPS, and promote its efficiency

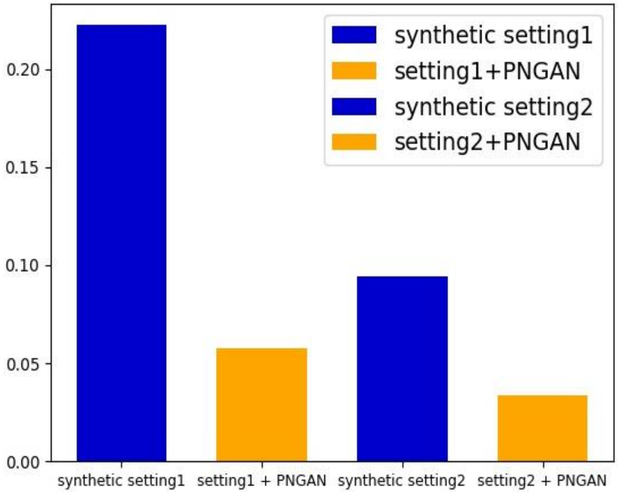
Experiments

Experiments

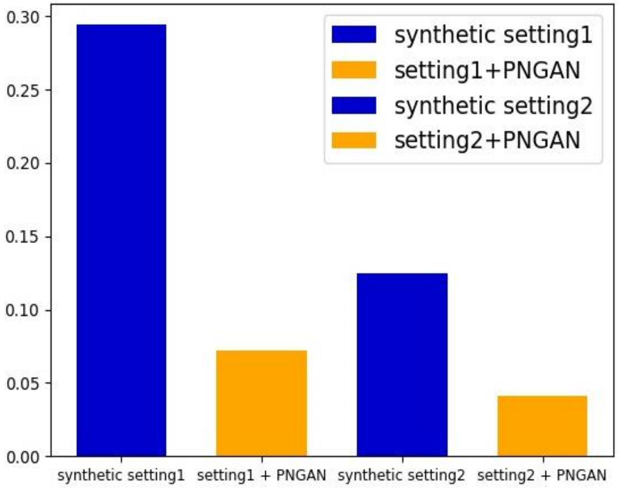
- Datasets
  - Real Noisy Image Pairs
    - SIDD : 320 image pairs for training, 1280 patch pairs for validation
    - DND : 50 image pairs, 1000 patch pairs cropped at size 512×512
    - PolyU : 40 real camera noisy-clean image pairs
    - Nam : real noisy image pairs of 11 static scenes
  - HD Clean Image
    - DIV2K, Flickr2K, BSD68, Kodak24, and Urban100

Experiments

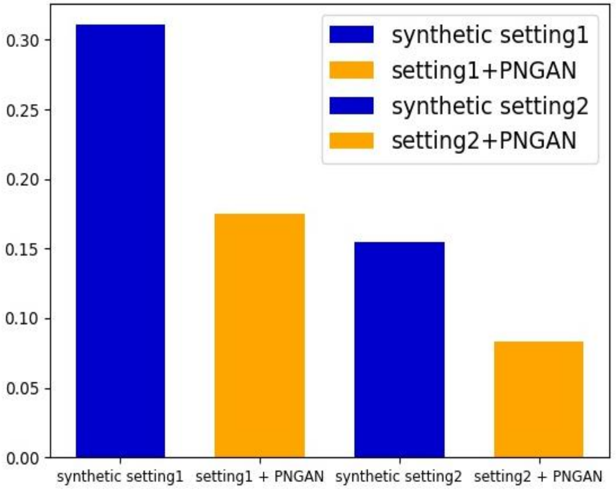
- MMD - Maximum Mean Discrepancy



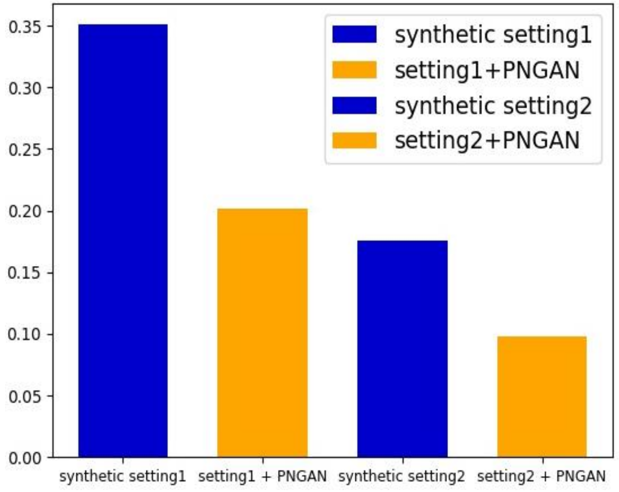
SIDD



DND



PolyU



Nam

- SIDD : Setting1 (74% ↓), Setting2 (64% ↓)
- DND : Setting1 (75% ↓), Setting2 (67% ↓)
- PolyU : Setting1 (44% ↓), Setting2 (46% ↓)
- Nam : Setting1 (43% ↓), Setting2 (44% ↓)

Experiments

- Comparisons with SOTA Methods

Methods	SIDD [43] PSNR ↑	SIDD [43] SSIM ↑	DND [44] PSNR ↑	DND [44] SSIM ↑	Methods	PolyU [37] PSNR ↑	PolyU [37] SSIM ↑	Nam [45] PSNR ↑	Nam [45] SSIM ↑
DnCNN-B [10]	23.66	0.583	32.43	0.790	RDN [12]	37.94	0.946	38.16	0.956
CBDNet [46]	33.28	0.868	38.06	0.942	FFDNet+ [17]	38.17	0.951	38.81	0.957
RIDNet [40]	38.71	0.914	39.26	0.953	TWSC [47]	38.68	0.958	38.96	0.962
AINDNet [48]	39.15	0.955	39.53	0.956	CBDNet [46]	38.74	0.961	39.08	0.969
VDN [49]	39.23	0.955	39.38	0.952	RIDNet [40]	38.86	0.962	39.20	0.973
CycleISP [20]	39.52	0.957	39.56	0.956	VDN [49]	39.04	0.965	39.68	0.976
MPRNet [50]	39.71	0.958	39.80	0.954	MPRNet [50]	39.07	0.969	39.41	0.974
MIRNet [51]	39.72	0.959	39.88	0.956	MIRNet [51]	39.18	0.973	39.57	0.979
RIDNet* (Ours)	**39.25**	**0.956**	**39.55**	**0.955**	**RIDNet* (Ours)**	**39.54**	**0.971**	**39.69**	**0.975**
MPRNet* (Ours)	**40.06**	**0.960**	**40.18**	**0.961**	**MPRNet* (Ours)**	**40.48**	**0.982**	**40.72**	**0.984**
MIRNet* (Ours)	**40.07**	**0.960**	**40.25**	**0.962**	**MIRNet* (Ours)**	**40.55**	**0.983**	**40.78**	**0.986**

- Improvements Achieved by Denoisers Finetuned with Fake Image Pairs

SIDD (0.35 dB ↑) DND (0.37 dB ↑)

PolyU (1.37 dB ↑) Nam (1.10 dB ↑)

Experiments

- Improvements of PNGAN on Different Datasets

Methods	SIDD [43] S1	SIDD [43] S1 + PNGAN		SIDD [43] S2	SIDD [43] S2+PNGAN		SIDD [43] Real	DF2K [53, 54] S1	DF2K [53, 54] S1 + PNGAN		DF2K [53, 54] S2	DF2K [53, 54] S2+PNGAN	
RIDNet	22.55	**37.92**	**(+15.37)**	36.13	**38.71**	**(+2.58)**	38.69	22.55	**32.10**	**(+9.55)**	33.98	**38.14**	**(+4.16)**
MPRNet	22.86	**38.52**	**(+15.66)**	36.52	**39.53**	**(+3.01)**	39.45	22.85	**32.82**	**(+9.97)**	34.19	**38.61**	**(+4.42)**
MIRNet	22.83	**38.76**	**(+15.93)**	36.55	**39.57**	**(+3.02)**	39.58	23.08	**32.34**	**(+9.26)**	34.26	**38.72**	**(+4.46)**

We use the fake noisy images generated from clean SIDD train and DF2K (DIV2K+Flicker2K) respectively to train denoisers from scratch.

SIDD : Setting1 (~ 15.65 dB ↑), Setting2 (~ 2.87 dB ↑)

DF2K : Setting1 (~ 9.59 dB ↑), Setting2 (~ 4.35 dB ↑)

Experiments

- Ablation Study

Methods	Baseline1	PNGAN Component + D_d		+ D		+ $\mathcal{L}_p$		Generator Architecture Baseline2	+ Multi-scale		+ SID		+ FCA	
RIDNet	14.54	35.37	(+20.83)	37.49	(+2.12)	37.92	(+0.43)	35.62	37.01	(+1.39)	37.23	(+0.22)	37.92	(+0.69)
MPRNet	14.25	36.26	(+22.01)	38.27	(+2.01)	38.52	(+0.25)	36.28	37.47	(+1.19)	37.86	(+0.39)	38.52	(+0.66)
MIRNet	13.57	36.15	(+22.58)	38.28	(+2.13)	38.76	(+0.48)	36.39	37.66	(+1.27)	37.89	(+0.23)	38.76	(+0.87)

Ablation Study of PNGAN Components

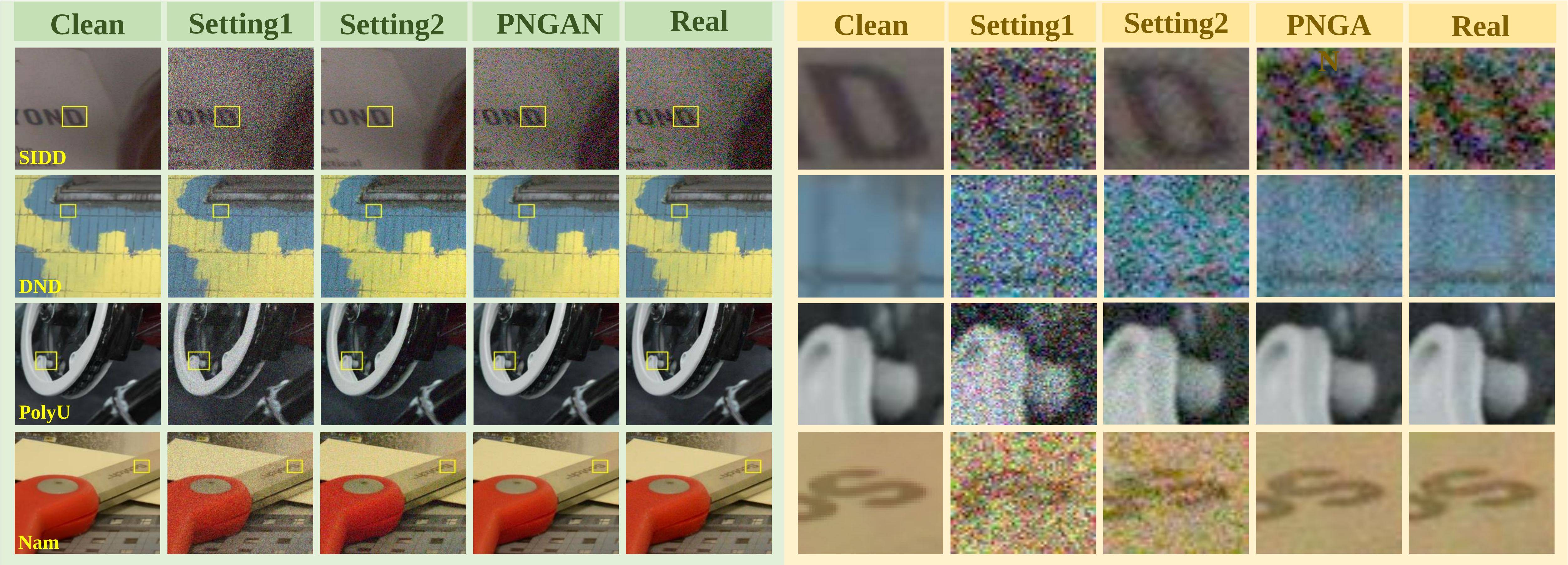
$+ D_d$: ($\sim$ 21.81 dB ↑), $+ D$: ($\sim$ 2.09 dB ↑), $+ L_p$: ($\sim$ 0.39 dB ↑)

Ablation Study of Generator Architecture

+ Multi-scale : ($\sim$ 1.28 dB ↑), + SID : ($\sim$ 0.28 dB ↑), + FCA: ($\sim$ 0.74 dB ↑)

Experiments

- Visual Examinations of Noisy Images



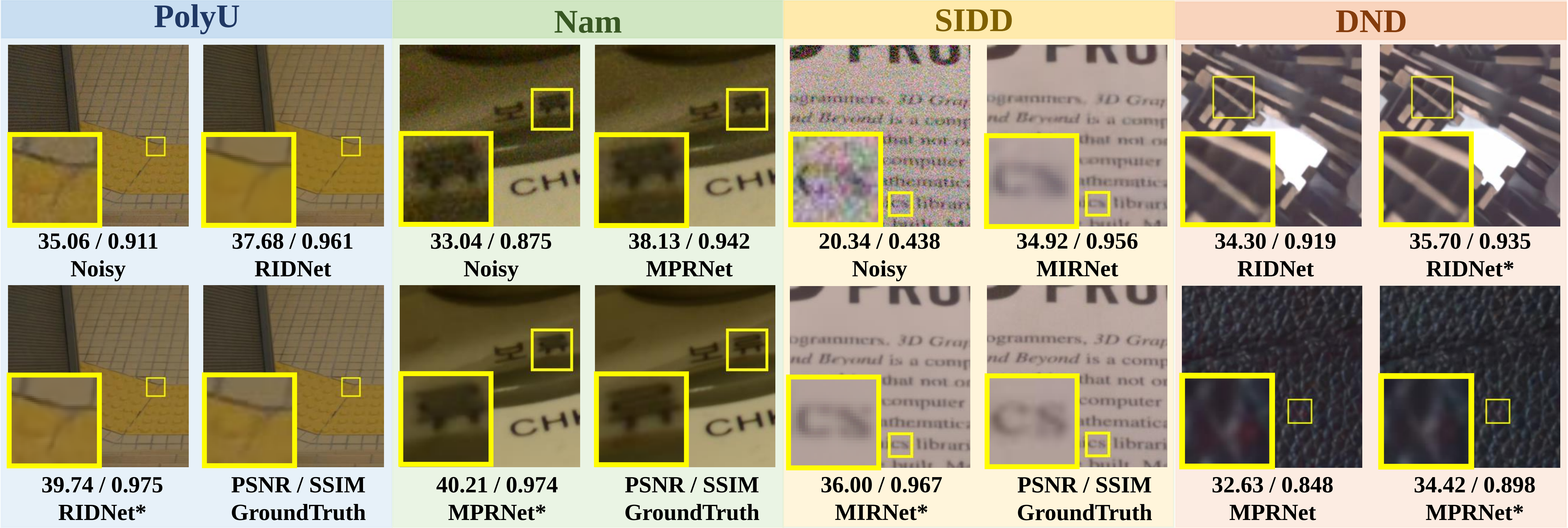
Setting1 : Gaussian Noise, fundamentally different from real noise

Setting2 : Poisson-Gaussian + Pseudo ISP, different in intensity and distribution

PNGAN : Can model spatio-chromatically correlated and non-Gaussian noise

Experiments

- Visual Comparisons of Denoised Images



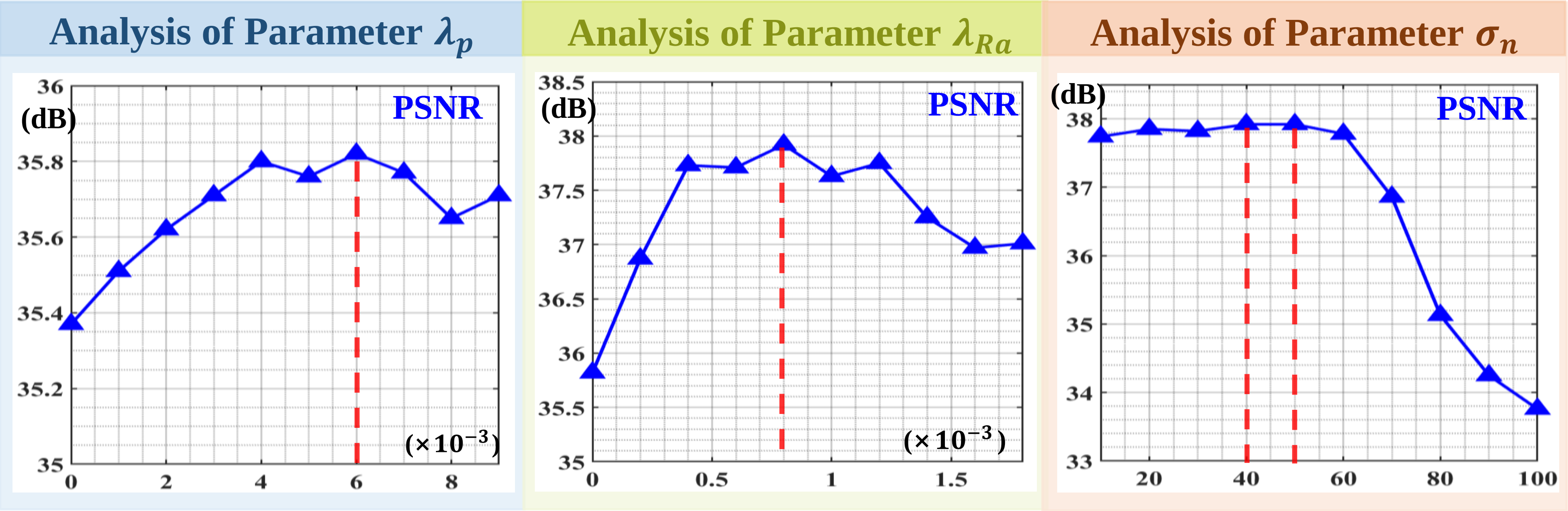
- Models finetuned with the generated data are capable of preserving the structural content, textural details, and spatial smoothness of the homogeneous regions.
- In contrast, original models either yield over-smooth images sacrificing fine textural details and structural content or introduce redundant blotchy texture and chroma artifacts.

Experiments

- Parameter Analysis

The optimal setting:

- $\lambda_p = 6 \times 10^{-3}$
- $\lambda_{Ra} = 8 \times 10^{-4}$
- $\sigma_n = 40$ or 50
- $q = 60\ \%$



q	SIDD [43] PSNR	SIDD [43] SSIM	PolyU [37] PSNR	PolyU [37] SSIM	Nam [45] PSNR	Nam [45] SSIM	Total PSNR	Total SSIM
None	38.71	0.914	38.86	0.962	39.20	0.973	38.76	0.929
0	**39.32**	**0.957**	38.01	0.949	38.34	0.958	38.92	0.955
20%	39.29	**0.957**	38.45	0.959	38.87	0.970	39.03	0.958
40%	39.28	0.956	39.02	0.966	39.26	0.973	39.20	0.959
60%	39.26	0.956	39.54	0.971	39.69	0.975	**39.35**	**0.961**
80%	39.23	0.955	39.56	**0.972**	39.72	**0.976**	39.33	0.960
100%	39.21	0.955	**39.57**	**0.972**	**39.73**	**0.976**	39.33	0.960

Thanks