NeRF in the Dark: High Dynamic Range View Synthesis from Noisy Raw Images

CVPR 2022 (Oral)

Ben Mildenhall, Peter Hedman, Ricardo Martin-Brualla,
Pratul P. Srinivasan, Jonathan T. Barron

First unit: Google Research

徐一伦

2022/10/16

Website: https://bmild.github.io/rawnerf/

Outline

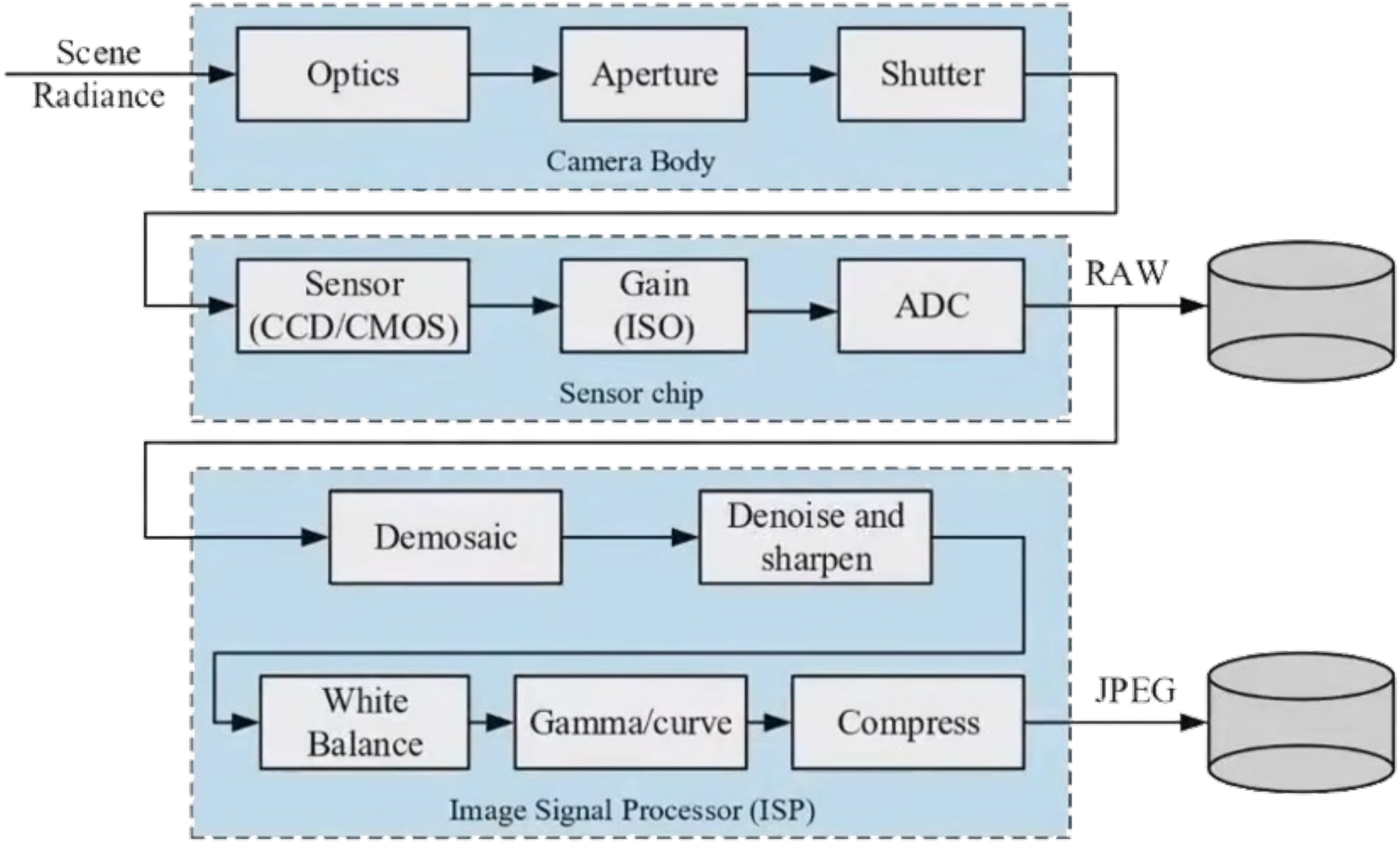
- Authorship
- Background
- Method
- Experiment
- Conclusion

Background-Image Processing

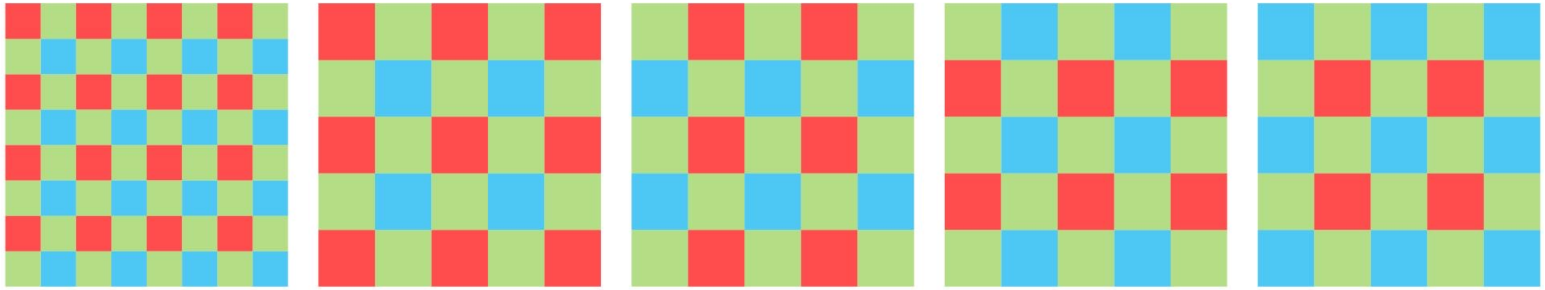
Image

Processing

Pipeline

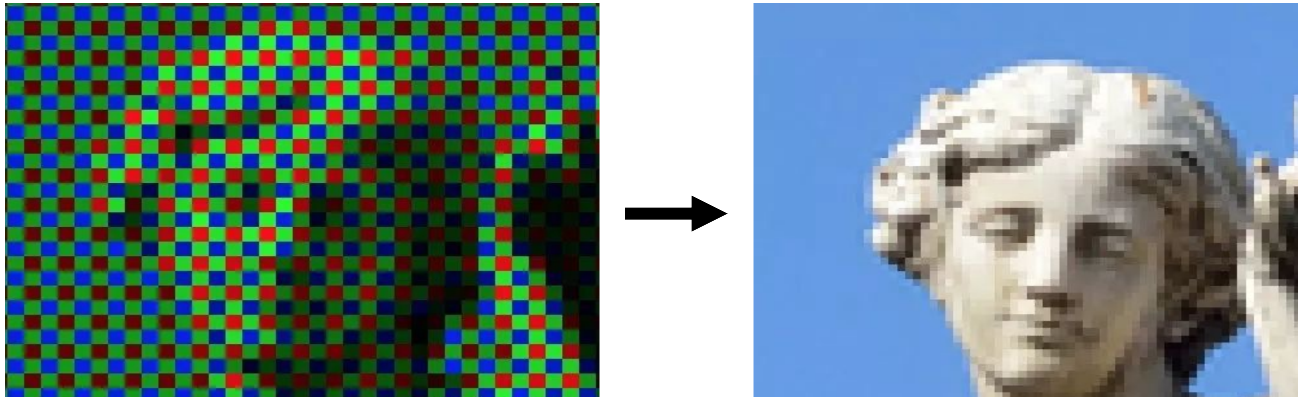


Demosaicing algorithm



Bayer Image | Pattern 1 | Pattern 2 | Pattern 3 | Pattern 4

Design different interpolation formulas for different patterns



The noise in raw image

Source of the noise

- "shot" noise: photon arrivals are a Poisson process
- "read" noise: noise in the readout circuitry

Model of the noise

$$z(x) = y(x) + \sigma\left(y(x)\right)\xi(x)$$

the distribution of the error is **zero-mean**.

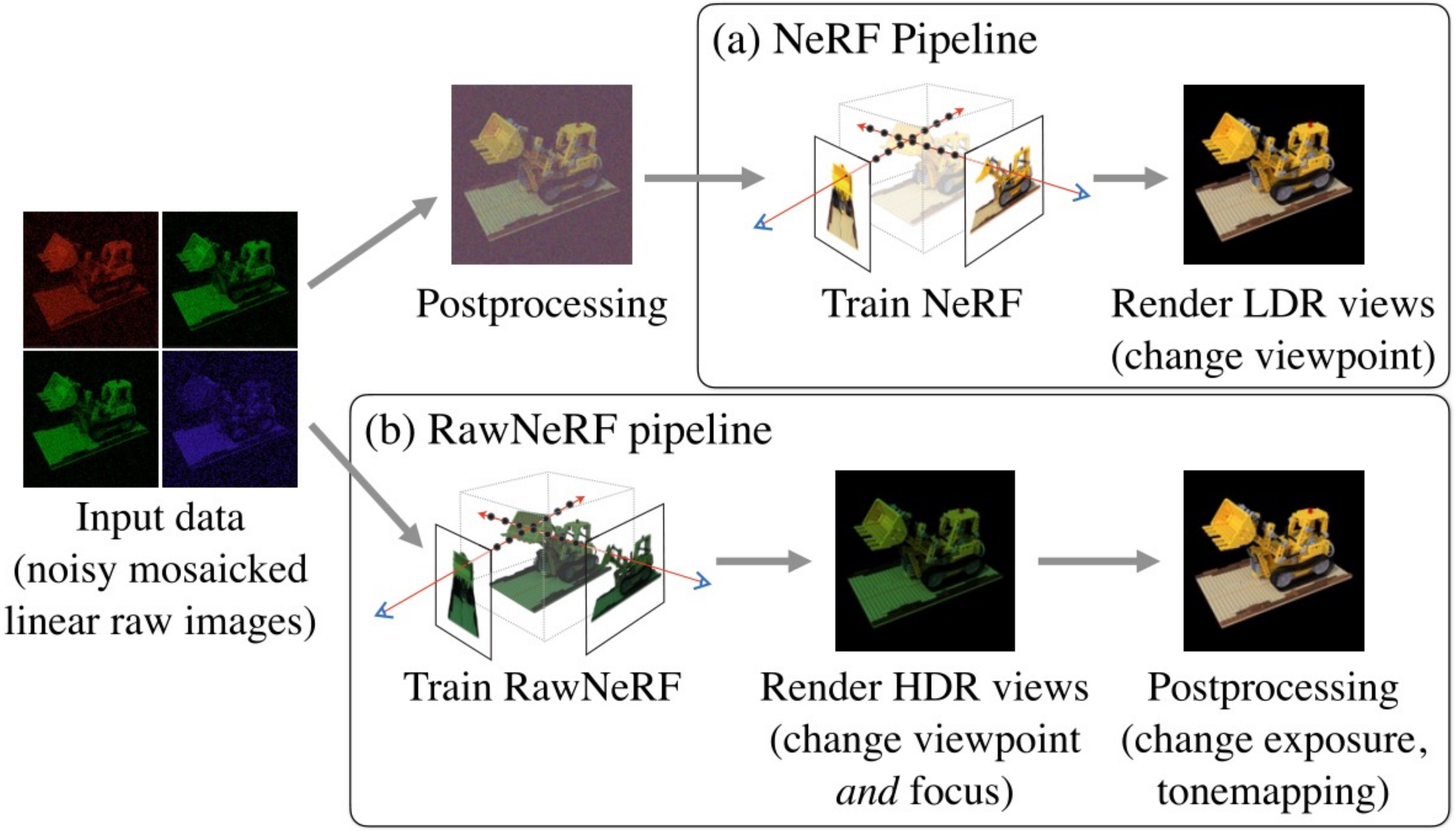
Alessandro et al. Practical poissonian-gaussian noise modeling and fitting for single-image raw-data, TIP 2008.

Outline

- Authorship
- Background
- Method
- Experiment
- Conclusion

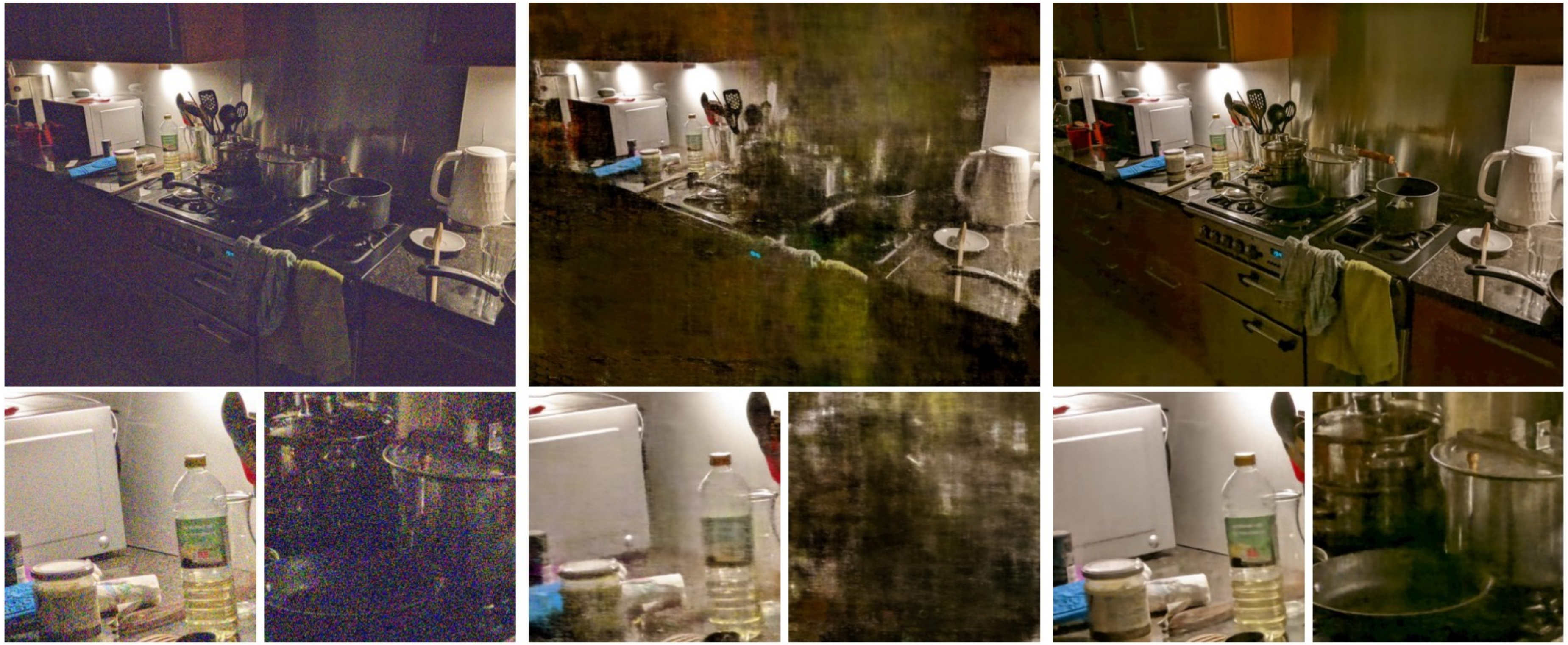
Method

RawReRF Pipeline



Method

The problem of L2 loss for training raw image



(a) Noisy raw test image (b) Trained w/ L2 loss (c) Trained w/ proposed loss

Muddy dark regions. Why?

Method

A simple way:

passing both the **rendered estimate** $\hat{y}$ and **noisy observed intensity** y

through a **tonemapping curve** ψ before the loss is applied

$$L_\psi(\hat{y}, y) = \sum_i (\psi(\hat{y}_i) - \psi(y_i))^2$$

The **nonlinear** tonemap will introduce **bias**

that change the noisy signal’s distribution.

Noise is not zero-mean.

Method

A better way:

To converge to an unbiased result

Using a locally valid linear approximation for the error term

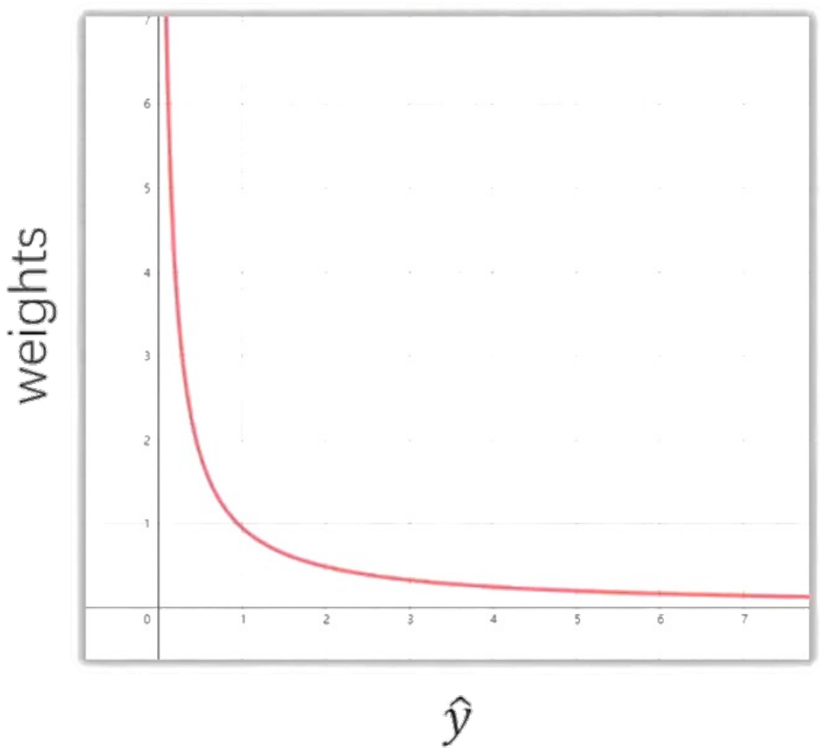
$$\psi(\hat{y}_i) - \psi(y_i) \approx \psi(\hat{y}_i) - (\psi(\hat{y}_i) + \psi'(\hat{y}_i)(y_i - \hat{y}_i))$$
$$= \psi'(\hat{y}_i)(\hat{y}_i - y_i)\,.$$

Note that author choose to linearize around $\hat{y}$ because, unlike the noisy observation y, $\hat{y}$ tends towards the true signal value $x_i = E[y_i]$ over the course of training.

Method

The weighted L2 loss: $$\tilde{L}_{\psi}(\hat{y}, y) = \sum_{i} \left[\psi'(\mathrm{sg}(\hat{y}_i))(\hat{y}_i - y_i)\right]^2$$

weight



$sg(\cdot)$ indicates a stop-gradient

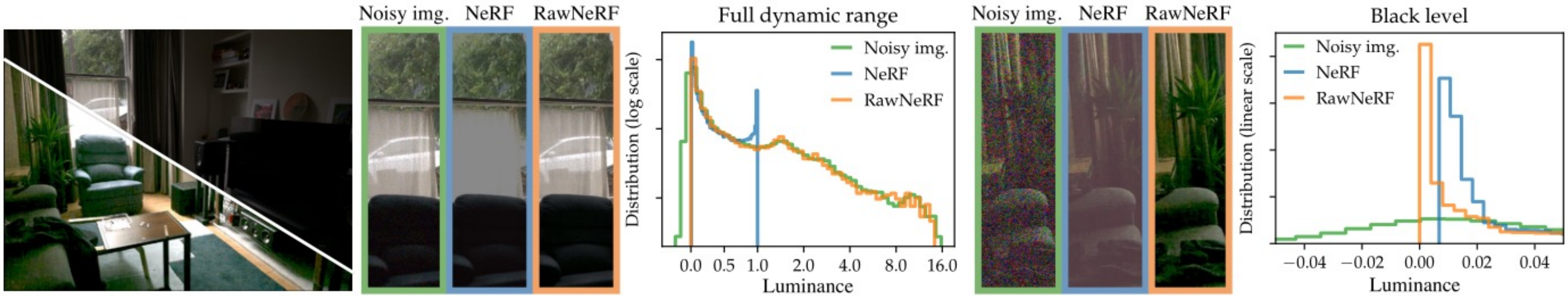
that treats its argument as an constant with zero derivative

Outline

- Authorship
- Background
- Method
- Experiment
- Conclusion

Experiment-denoising

Failure modes of NeRF on a daytime indoor scene



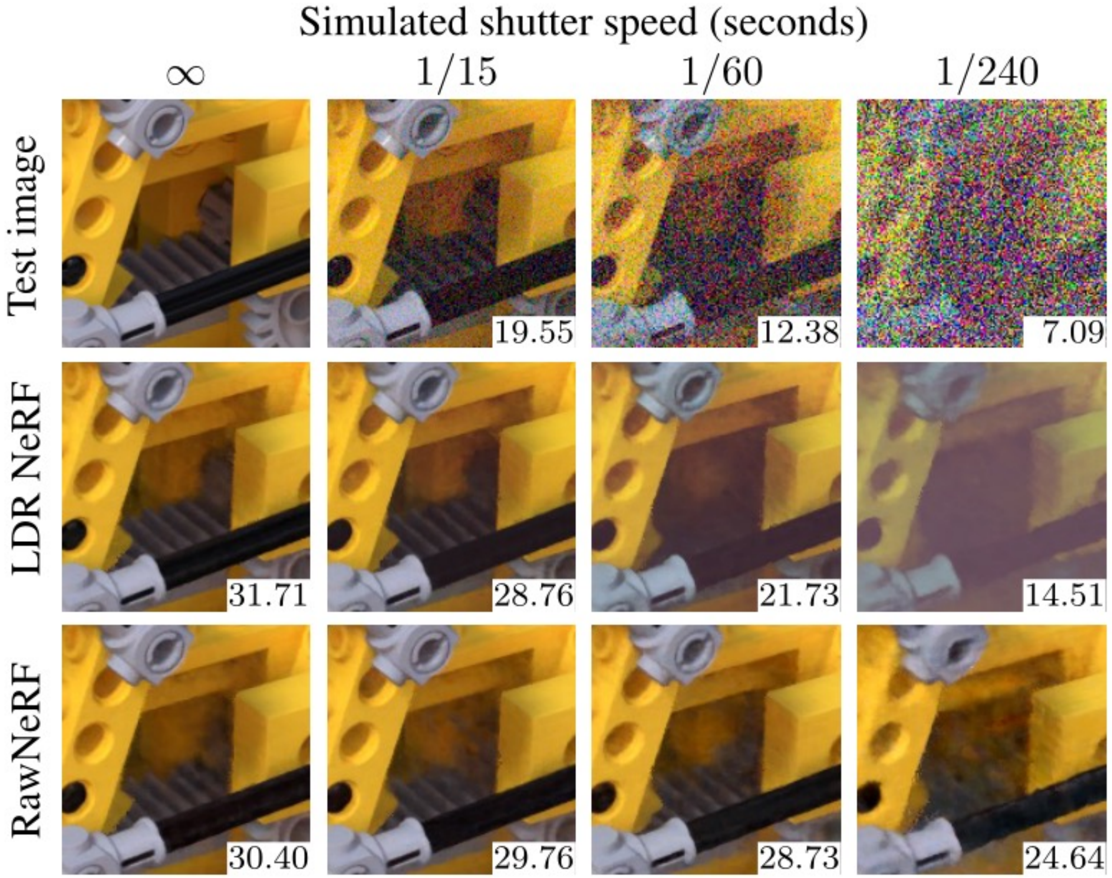
(a) Full RawNeRF output (b) Bright region: RawNeRF preserves highlights (c) Dark region: RawNeRF avoids bias

Experiment-denoising

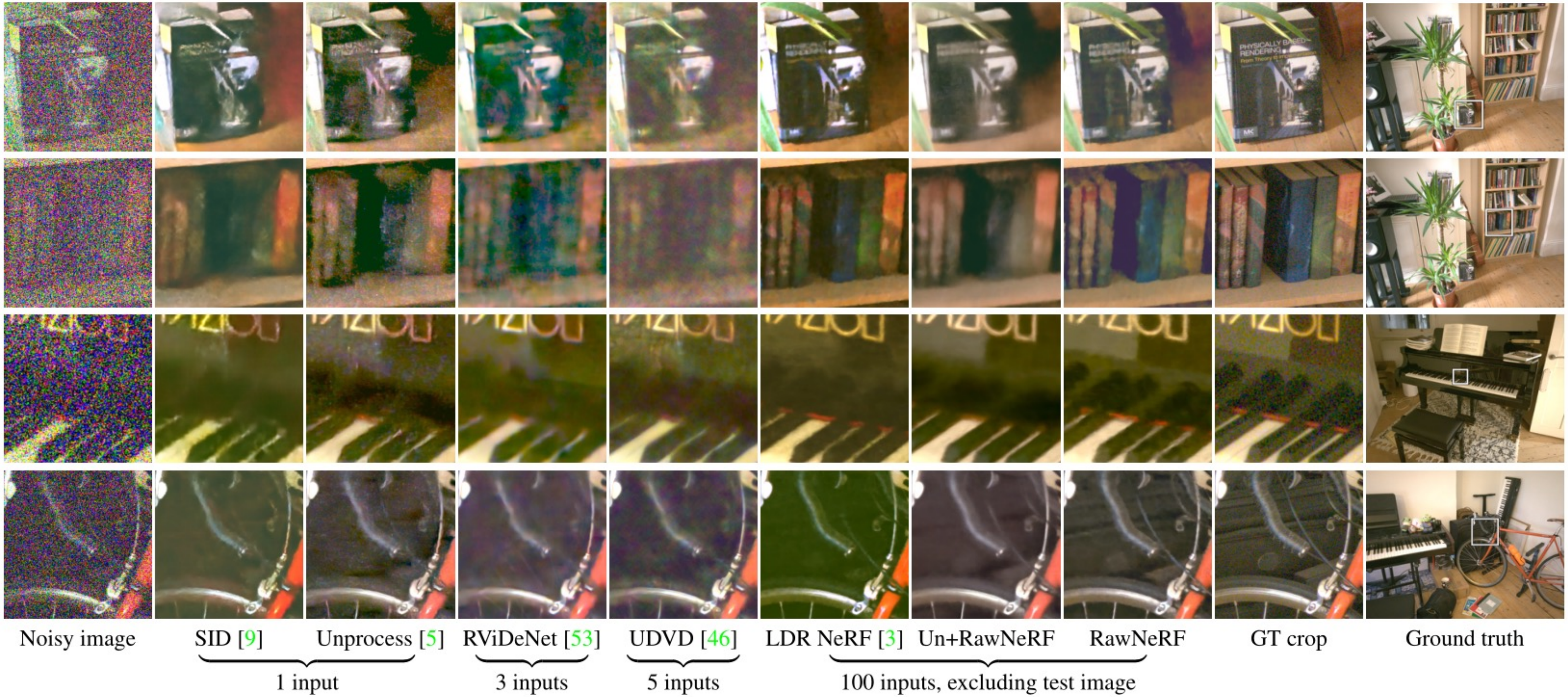
ablation study

on synthetic scene

Method	Simulated shutter speed (seconds) ∞	1/7	1/15	1/30	1/60	1/120	1/240
Noisy input	-	23.33	19.65	16.03	12.51	9.40	7.18
LDR NeRF	**33.16**	31.25	29.14	26.10	22.31	18.27	14.87
RawNeRF	32.15	**32.11**	**31.94**	**31.59**	**30.94**	**29.69**	**27.73**



Experiment-denoising



Experiment-denoising

Method	Num. inputs	Raw PSNR↑	Affine-aligned sRGB PSNR↑	SSIM↑	LPIPS↓
Noisy input	-	54.38	10.24	0.035	0.733
SID [9]	1	-	21.62	0.525	0.547
Unprocess [5]	1	70.80	23.02	0.491	0.489
RViDeNet [53]	3	68.29	22.20	0.516	0.545
UDVD [46]	5	70.68	22.75	0.514	0.507
LDR NeRF [3]	$N-1$	-	19.43	0.518	0.544
Un+RawNeRF	$N-1$	67.99	23.35	0.531	0.507
RawNeRF	$N-1$	67.20	23.53	0.536	0.501

Every deep denoiser uses pretrained model weights

Experiment-application

Synthetic defocus, modifying exposure and tonemapping

Video website: https://bmild.github.io/rawnerf/img/candle_focus_crop.mp4

Outline

- Authorship
- Background
- Method
- Experiment
- Conclusion

Conclusion

Contribution:

- Proposing a method for training RawNeRF directly on raw images that can handle high dynamic range scenes as well as noisy inputs captured in the dark.

Future work:

- Jointly optimize RawNeRF and calculate the input camera poses
- Training on raw images with variable exposure

Thanks for watching.

徐一伦

yilunxu_buaa@163.com