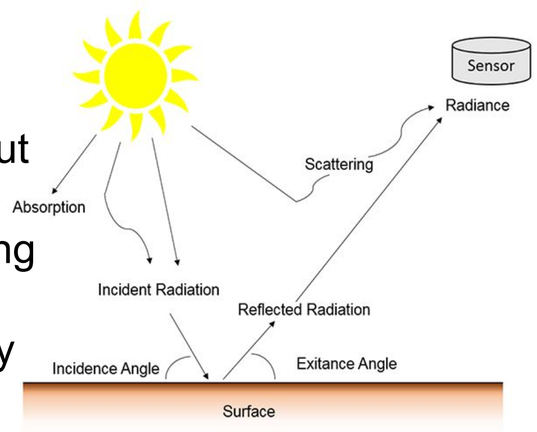
ERD: Extended RAW-Diffusion Framework for De-rendering sRGB Images

Jiaqi Shang, Yifan Qu
Department of Computer Science, University of Toronto

Introduction

Background

- **RAW** data preserves **true** scene radiance but is lost after ISP processing, making **inverse** ISP essential for many low-level imaging tasks.



Motivation

- Existing methods rely on **known ISP** steps, or specialized capture conditions, **limiting generalization.**
- Old RAW Diffusion method [1] improves reconstruction but still needs **per-sensor** training and lacks metadata for fairness and bias evaluation.

Contribution

- We propose a **camera-agnostic** inverse ISP framework that reconstructs RAW images from **any sRGB** input while improving generalization and accuracy across arbitrary sensors.

Dataset

- Adobe FiveK (Nikon and Canon)
- RAW-NOD (Nikon and Sony)

Related Work

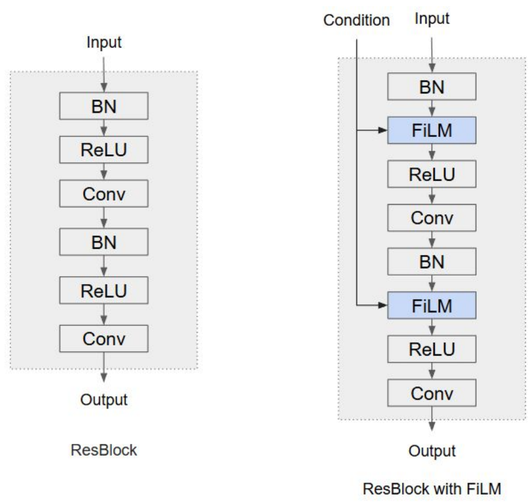
- **Invertible ISP designs:** Methods like InvISP [2] redesign the forward ISP to be mathematically invertible, but only apply when the ISP itself can be modified.
- **Learning-based sampling and reconstruction**: Works such as Nam et al. [3] use superpixel-based sampling with multi-phase training but still assume known ISP pipelines and struggle with cross-camera accuracy.
- **Generative models:** RAW Diffusion [1] leverages a UNet with multi-resolution sRGB guidance but requires training a separate model per sensor and lacks metadata for fairness/bias evaluation, constraining its practicality.

References

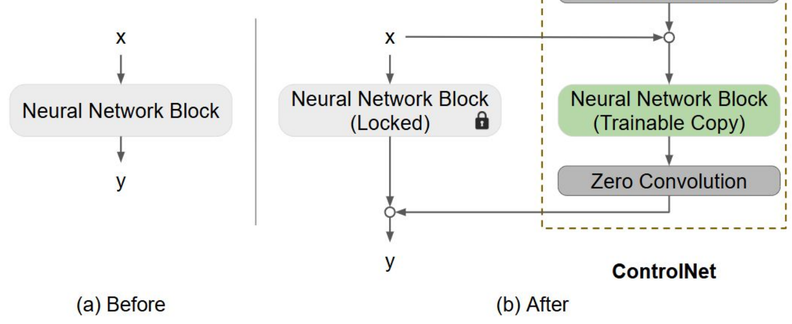
[1] Reinders et al. "Raw-diffusion: Rgb-guided diffusion models for high-fidelity raw image generation." 2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). IEEE, 2025.
[2] Xing et al. "Invertible image signal processing." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021.
[3] Nam et al. "Learning srgb-to-raw-rgb de-rendering with content-aware metadata." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2022.

Methodology

Global Conditioning with FiLM

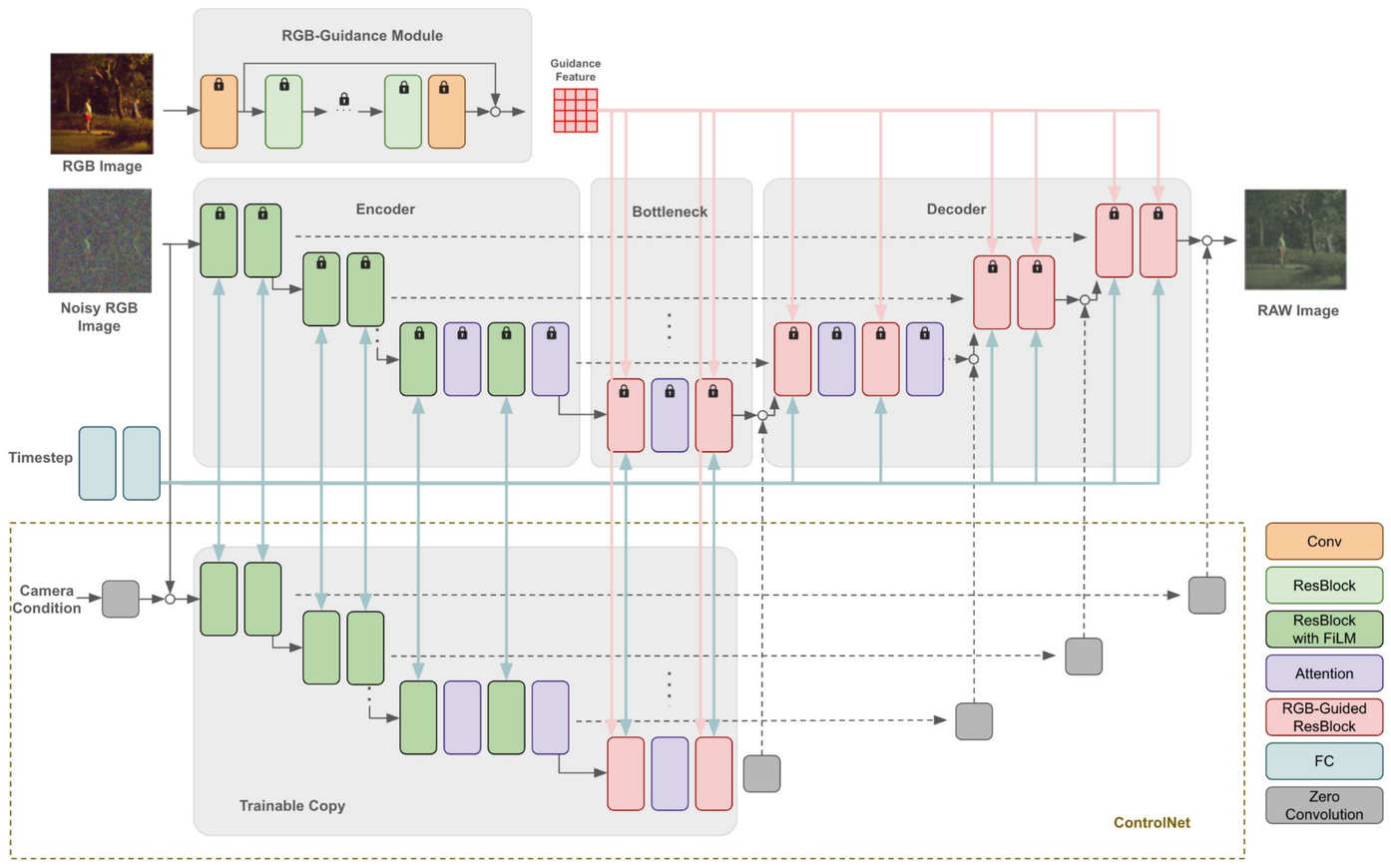


ControlNet Branch



Computationally Efficient ControlNet Integration

- Reuse the RAW-Diffusion encoder and bottleneck



Experimental Results

- We trained our mode on A100 GPU, with 10000 steps using AdamW. Initial learning rate is 0.0001, linearly decreasing to zero, and a batch size of 4.

- Our model performs comparably to the original framework while achieving **cross-camera** generalization.
- For future work, we aim to extend this generalization capability to a broader range of cameras.

	FiveK Canon		FiveK Nikon	
	PSNR	SSIM	PSNR	SSIM
RD (Canon)	**33.45**	**0.8981**	22.37	0.7549
RD (Nikon)	24.69	0.7465	28.86	0.8173
RD+Film (Canon)	31.73	0.8628	21.84	0.7241
RD+Film (Nikon)	24.50	0.7397	29.62	0.8357
RD+CN	**33.06**	**0.8914**	**30.87**	**0.8753**
RD+CN+Film	31.78	0.8548	30.37	0.8679