CVPR 2021 Tutorial:
Normalizing Flows and Invertible Neural Networks in Computer Vision

Normalizing Flows for Computer Vision: Wavelet Flow and Noise Flow

Marcus A. Brubaker

VECTOR INSTITUTE

LASSONDE SCHOOL OF ENGINEERING

YORK UNIVERSITÉ UNIVERSITY

Flows and INNs in Computer Vision: CVPR 2021 Edition

- ***DeFlow: Learning Complex Image Degradations From Unpaired Data With Conditional Flows*** by Valentin Wolf, Andreas Lugmayr, Martin Danelljan, Luc Van Gool, Radu Timofte
- ***ArtFlow: Unbiased Image Style Transfer via Reversible Neural Flows*** by Jie An, Siyu Huang, Yibing Song, Dejing Dou, Wei Liu, Jiebo Luo
- ***iVPF: Numerical Invertible Volume Preserving Flow for Efficient Lossless Compression*** by Shifeng Zhang, Chen Zhang, Ning Kang, Zhenguo Li.
- ***Generative Classifiers as a Basis for Trustworthy Image Classification*** by Radek Mackowiak, Lynton Ardizzone, Ullrich Kothe, Carsten Rother
- ***Flow-Based Kernel Prior With Application to Blind Super-Resolution*** by Jingyun Liang, Kai Zhang, Shuhang Gu, Luc Van Gool, Radu Timofte
- ***Autoregressive Stylized Motion Synthesis With Generative Flow*** by Yu-Hui Wen, Zhipeng Yang, Hongbo Fu, Lin Gao, Yanan Sun, Yong-Jin Liu
- ***Mol2Image: Improved Conditional Flow Models for Molecule to Image Synthesis*** by Karren Yang, Samuel Goldman, Wengong Jin, Alex X. Lu, Regina Barzilay, Tommi Jaakkola, Caroline Uhler
- ***Invertible Image Signal Processing*** by Yazhou Xing, Zian Qian, Qifeng Chen
- ***Invertible Denoising Network: A Light Solution for Real Noise Removal*** by Yang Liu, Zhenyue Qin, Saeed Anwar, Pan Ji, Dongwoo Kim, Sabrina Caldwell, Tom Gedeon
- ***Large-Capacity Image Steganography Based on Invertible Neural Networks*** by Shao-Ping Lu, Rong Wang, Tao Zhong, Paul L. Rosin
- ***Quality-Agnostic Image Recognition via Invertible Decoder*** by Insoo Kim, Seungju Han, Ji-won Baek, Seong-Jin Park, Jae-Joon Han, Jinwoo Shin
- ***Neural Parts: Learning Expressive 3D Shape Abstractions With Invertible Neural Networks*** by Despoina Paschalidou, Angelos Katharopoulos, Andreas Geiger, Sanja Fidler

Noise Flow:
Noise Modelling with Conditional Normalizing Flows

Abdelrahman Abdelhamed

Michael S. Brown

LASSONDE SCHOOL OF ENGINEERING

YORK UNIVERSITÉ UNIVERSITY

BOREALIS AI

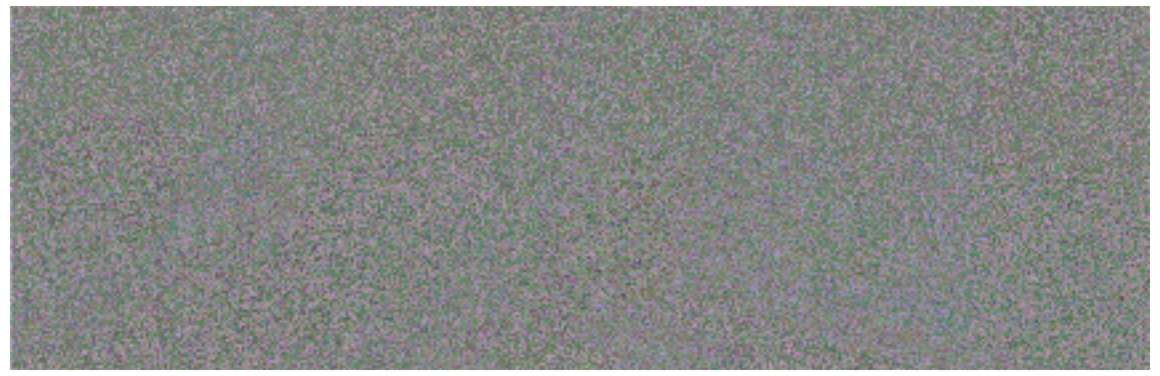
Camera Noise

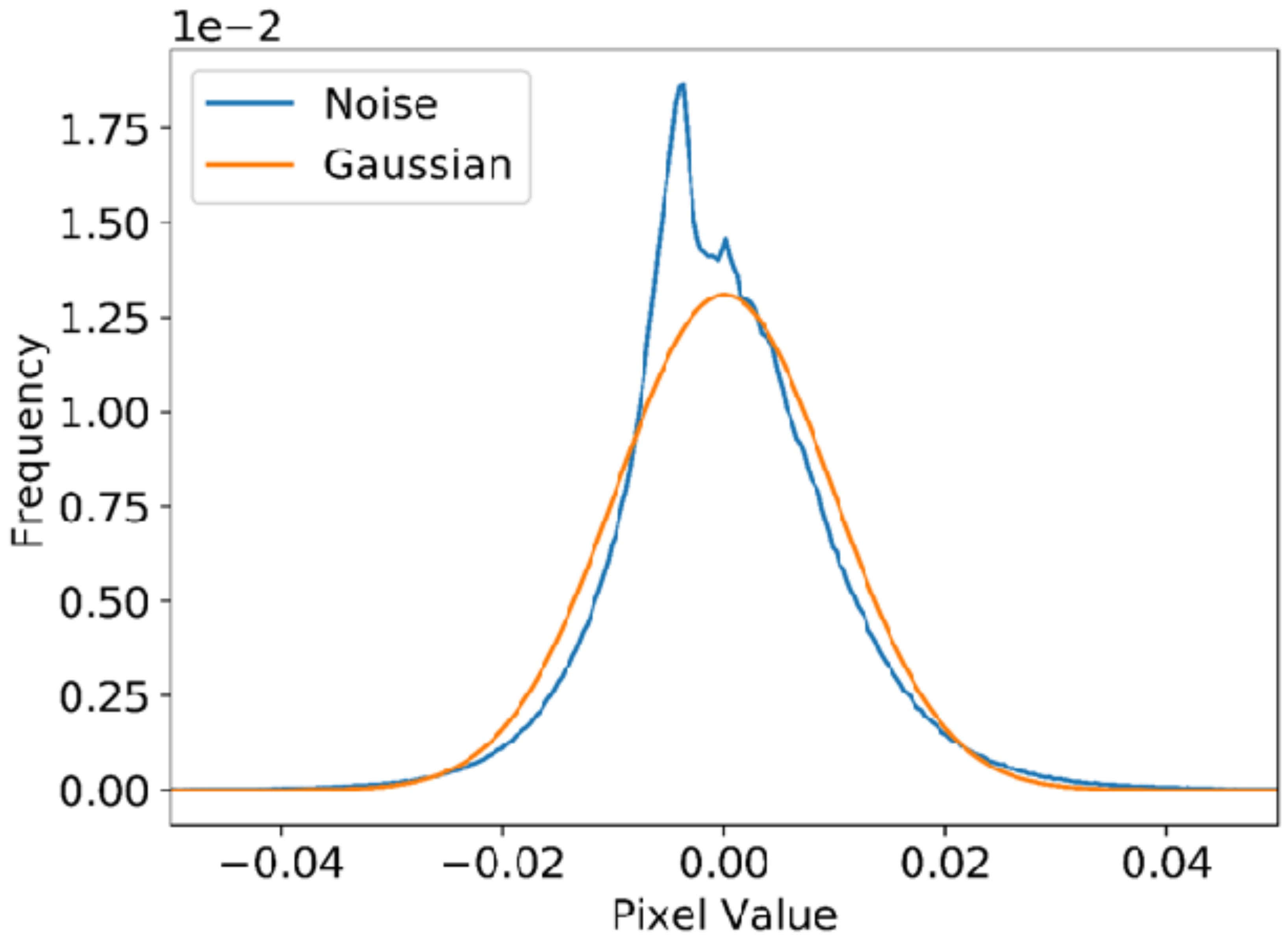
Image + Noise

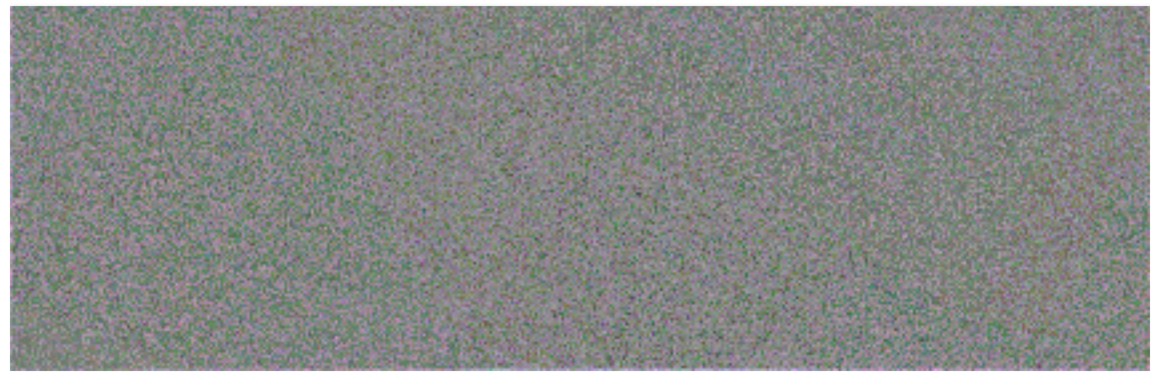
Camera: Pixel
ISO: 800
Exposure: 1/350 s

Camera Noise

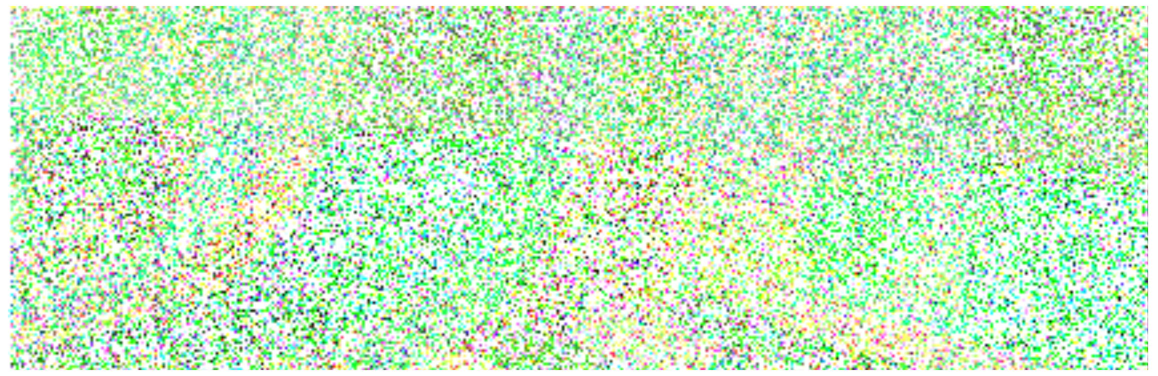
Camera Noise

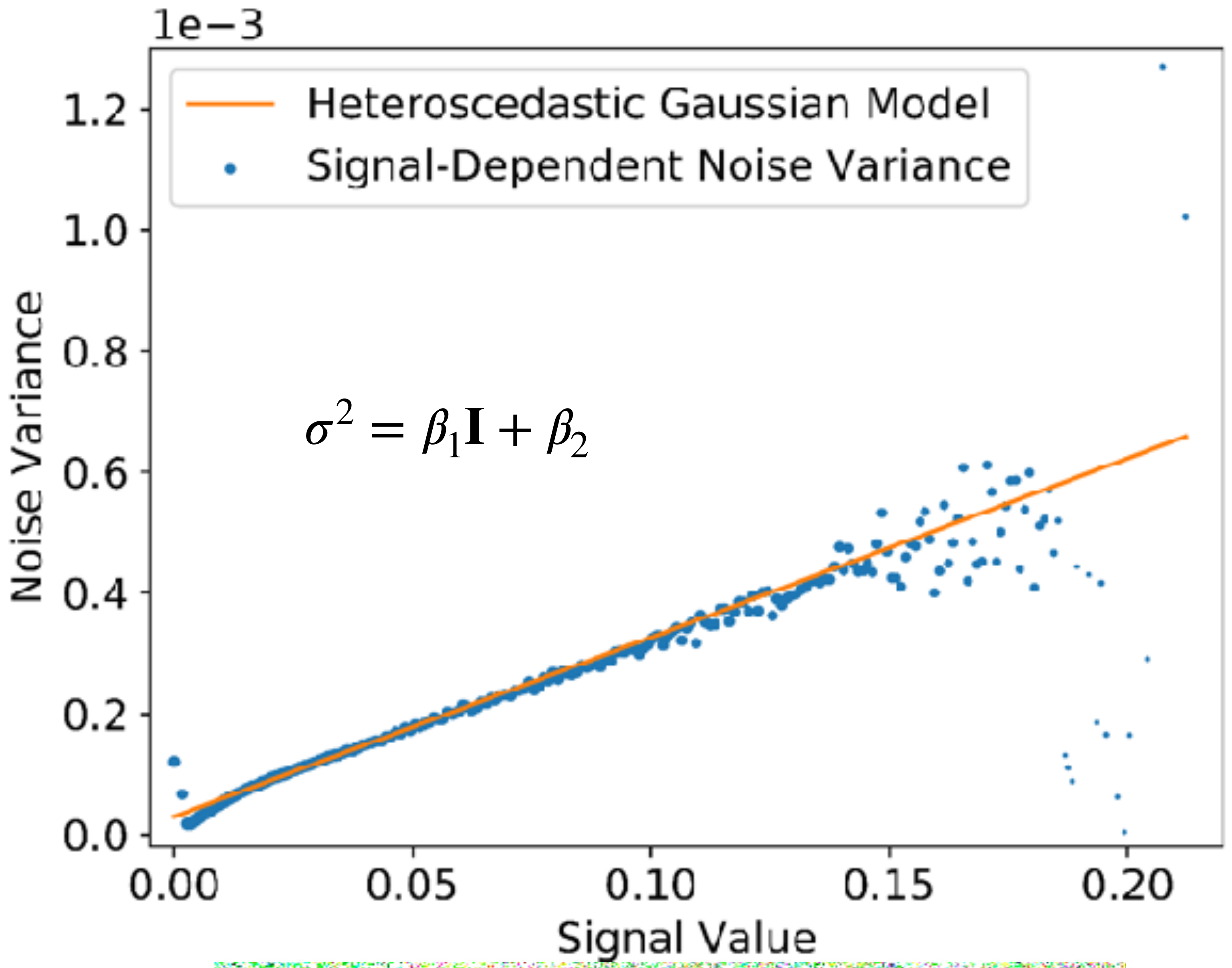


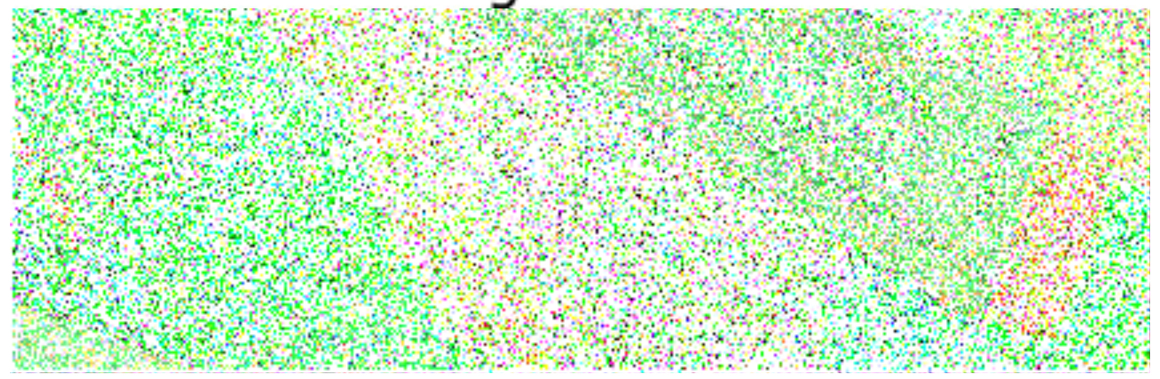




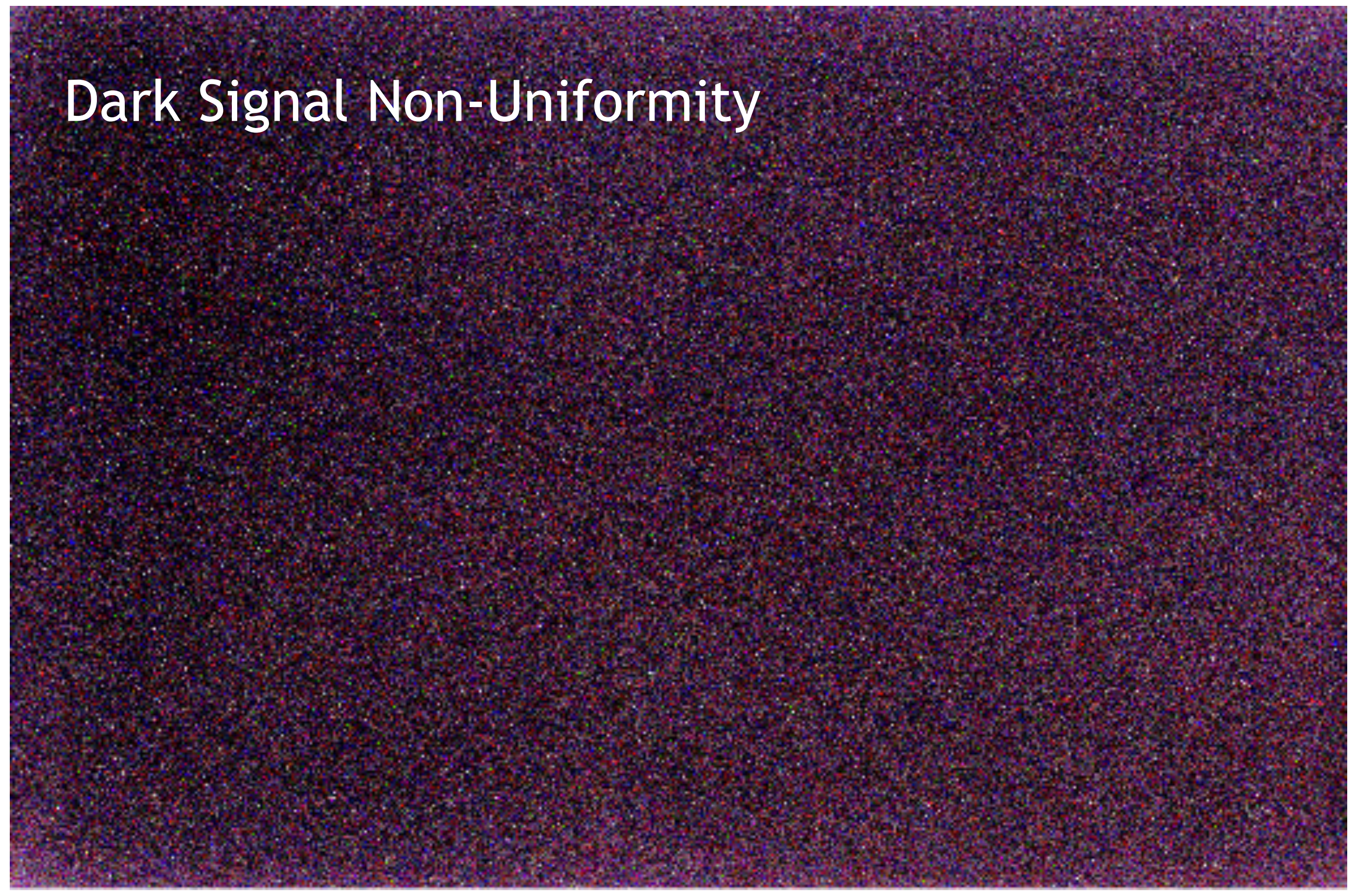
Camera Noise

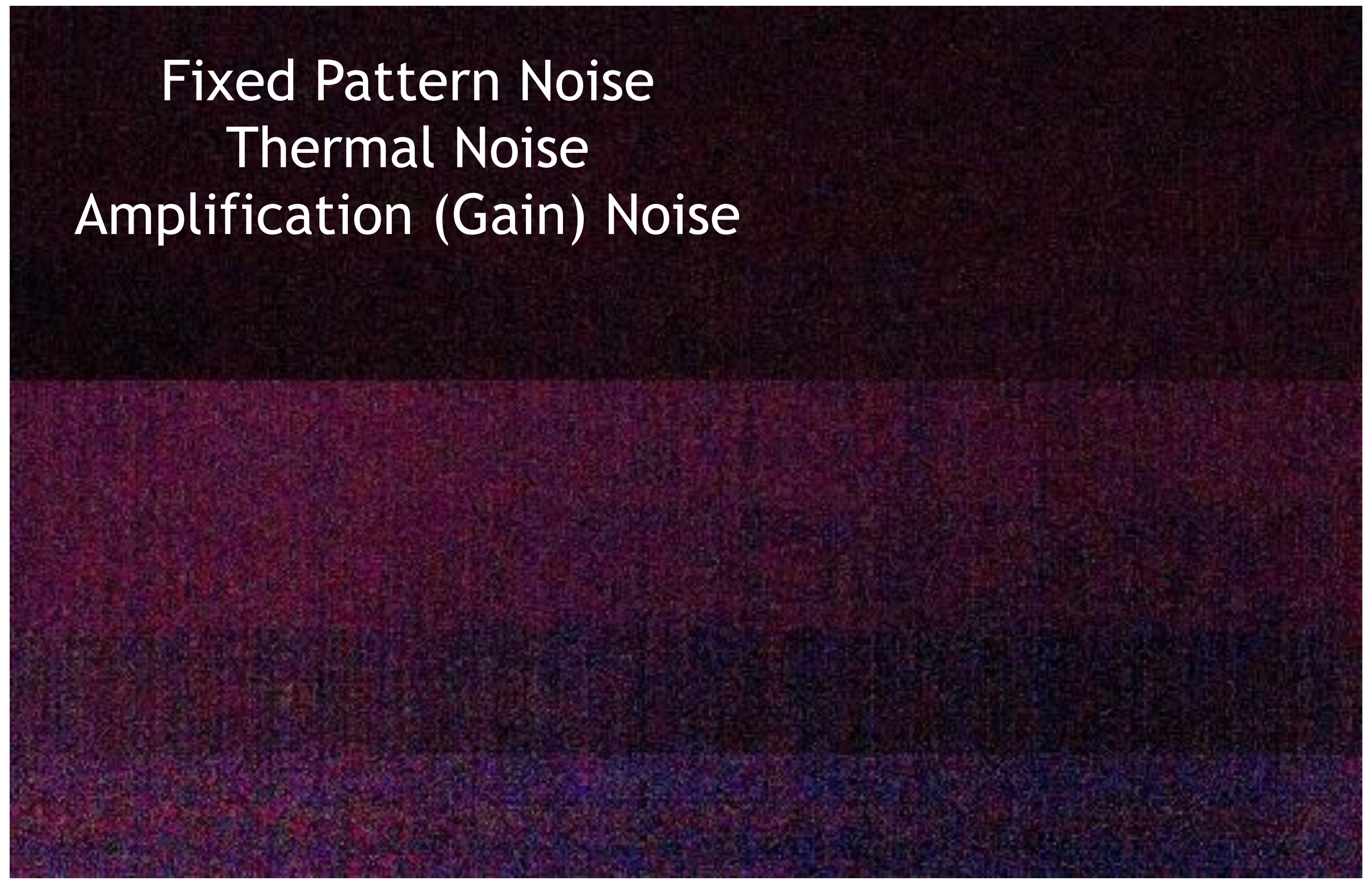


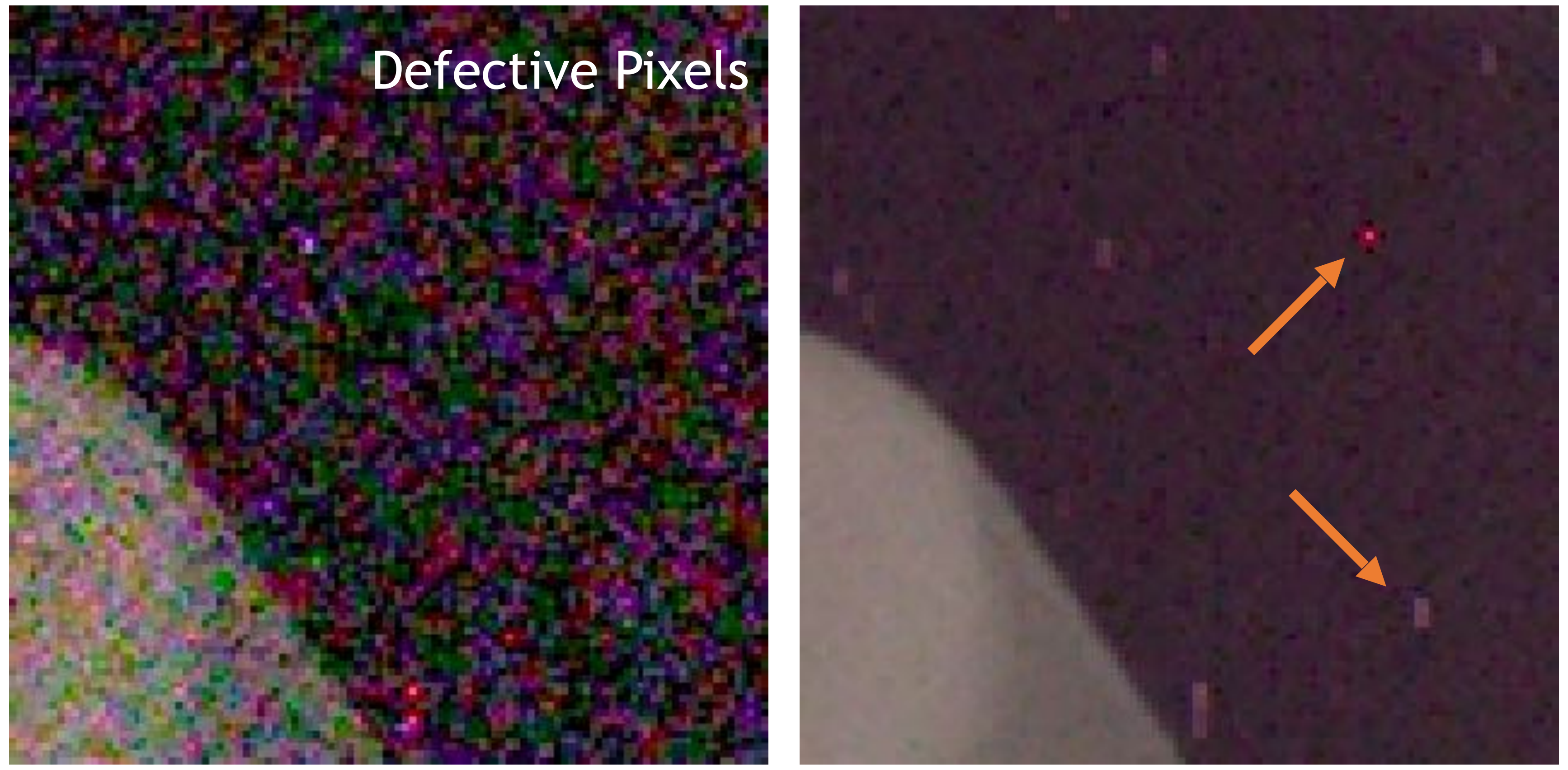


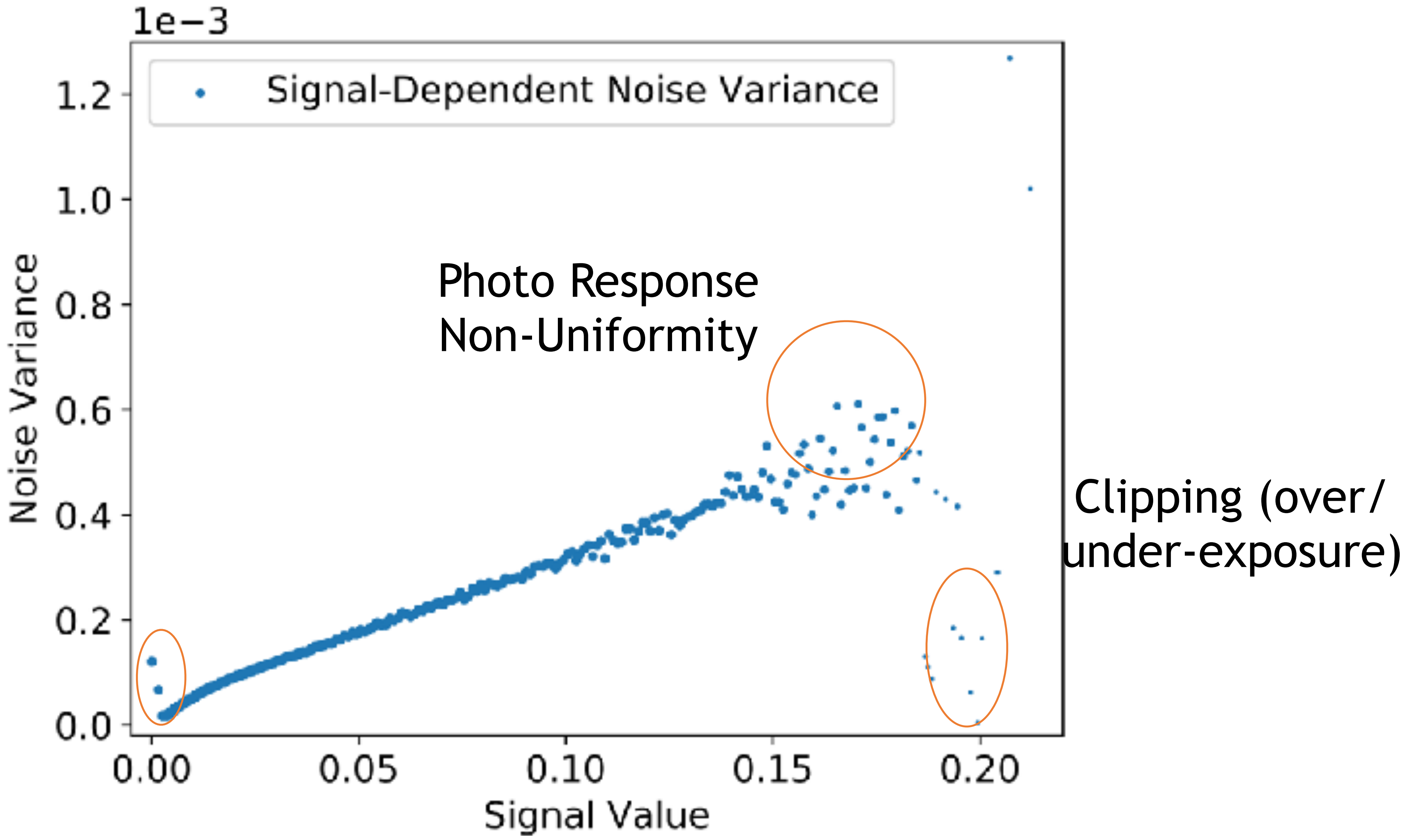


Camera Noise

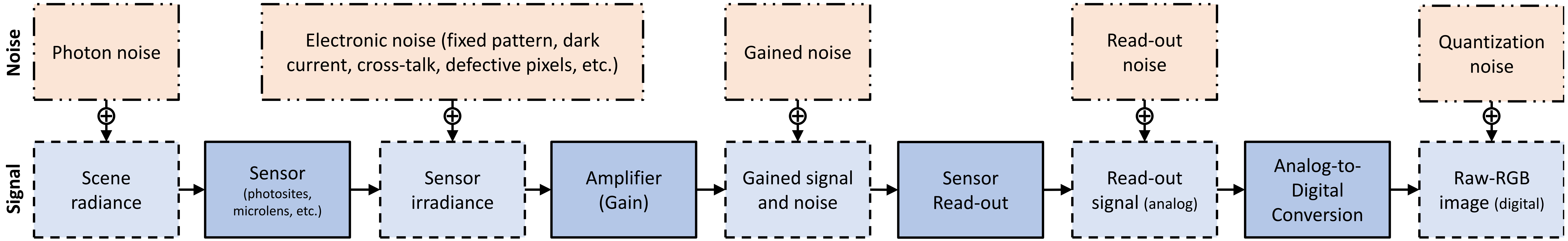








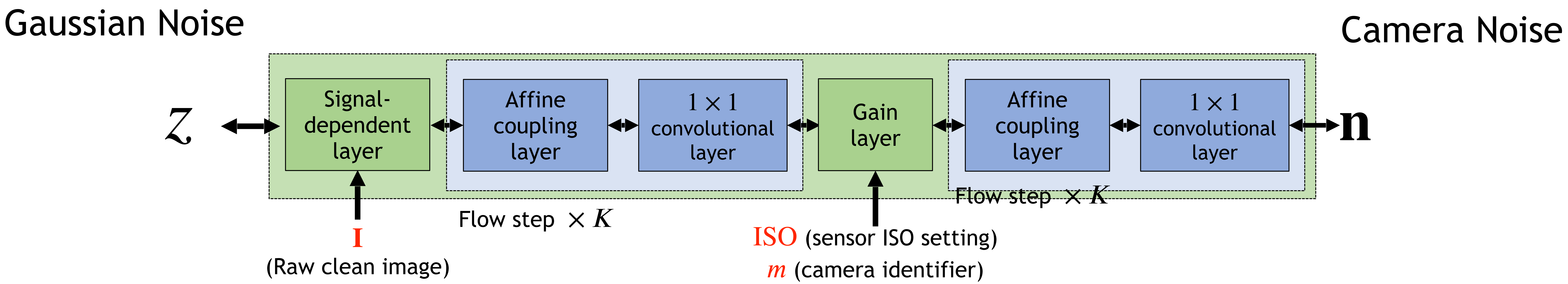
Camera Noise



Idea:

learn a convenient, compact model of camera noise
which exploits this knowledge using normalizing flows

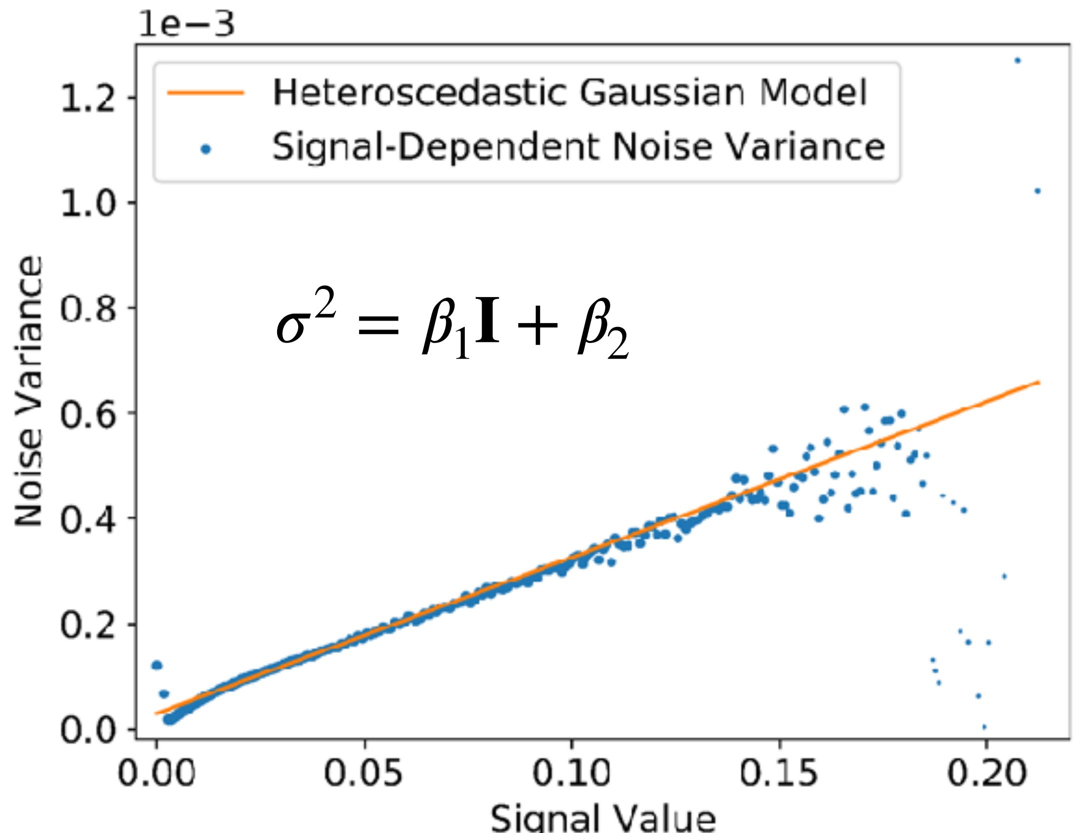
Noise Flow



$$g(z) = \mathbf{s} \odot z$$

$$\mathbf{s} = \left(\beta_1 \mathbf{I} + \beta_2\right)^{\frac{1}{2}}$$

$$g(z) = \gamma(\mathrm{ISO}, m) \odot z$$



$$\gamma(\mathrm{ISO}, m) = \psi_m \times u(\mathrm{ISO}) \times \mathrm{ISO}$$

u: scaling factor for ISO)

ψ_m: scaling factor for camera m

Smartphone Image Denoising Dataset (SIDD)

A High-Quality Denoising Dataset for Smartphone Cameras

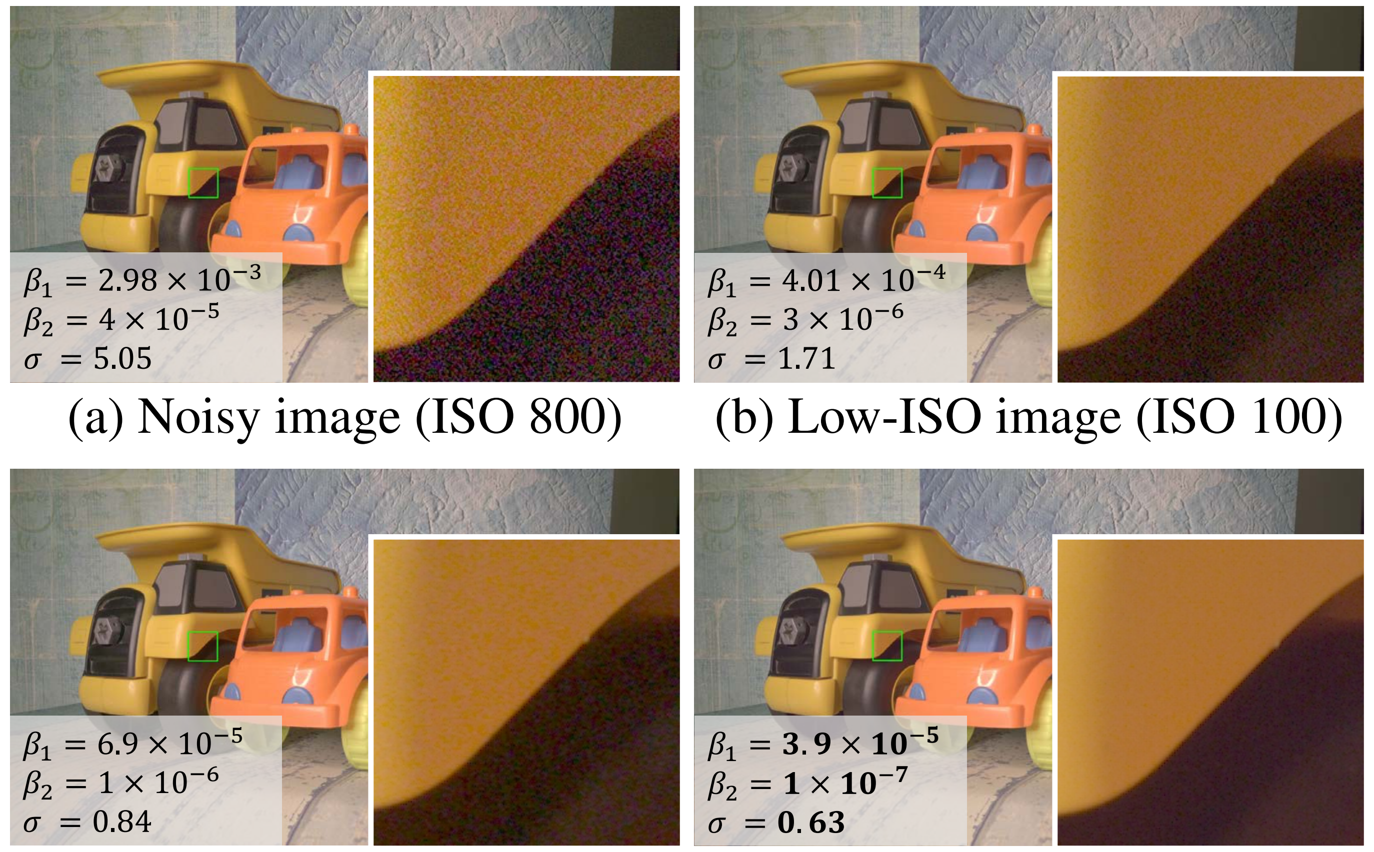
Abdelrahman Abdelhamed
York University
kamel@eecs.yorku.ca

Stephen Lin
Microsoft Research
stevelin@microsoft.com

Michael S. Brown
York University
mbrown@eecs.yorku.ca

Abstract

The last decade has seen an astronomical shift from imaging with DSLR and point-and-shoot cameras to imaging with smartphone cameras. Due to the small aperture and sensor size, smartphone images have notably more noise than their DSLR counterparts. While denoising for smartphone images is an active research area, the research community currently lacks a denoising image dataset representative of real noisy images from smartphone cameras with high-quality ground truth. We address this issue in this paper with the following contributions. We propose a systematic procedure for estimating ground truth for noisy images that can be used to benchmark denoising performance for smartphone cameras. Using this procedure, we have captured a dataset – the Smartphone Image Denoising Dataset (SIDD) *– of ~30,000 noisy images from 10 scenes under different lighting conditions using five representative*



(a) Noisy image (ISO 800) (b) Low-ISO image (ISO 100)
(c) Ground truth using [25] (d) Our ground truth

Figure 1: An example scene imaged with an LG *G4* smartphone camera: (a) a high-ISO noisy image; (b) same scene captured with low ISO – this type of image is often used as ground truth for (a); (c) ground truth estimated by [25]; (d) our ground truth. Noise estimates (β_1 and β_2 for noise level

CVPR 2018

Results on SIDD

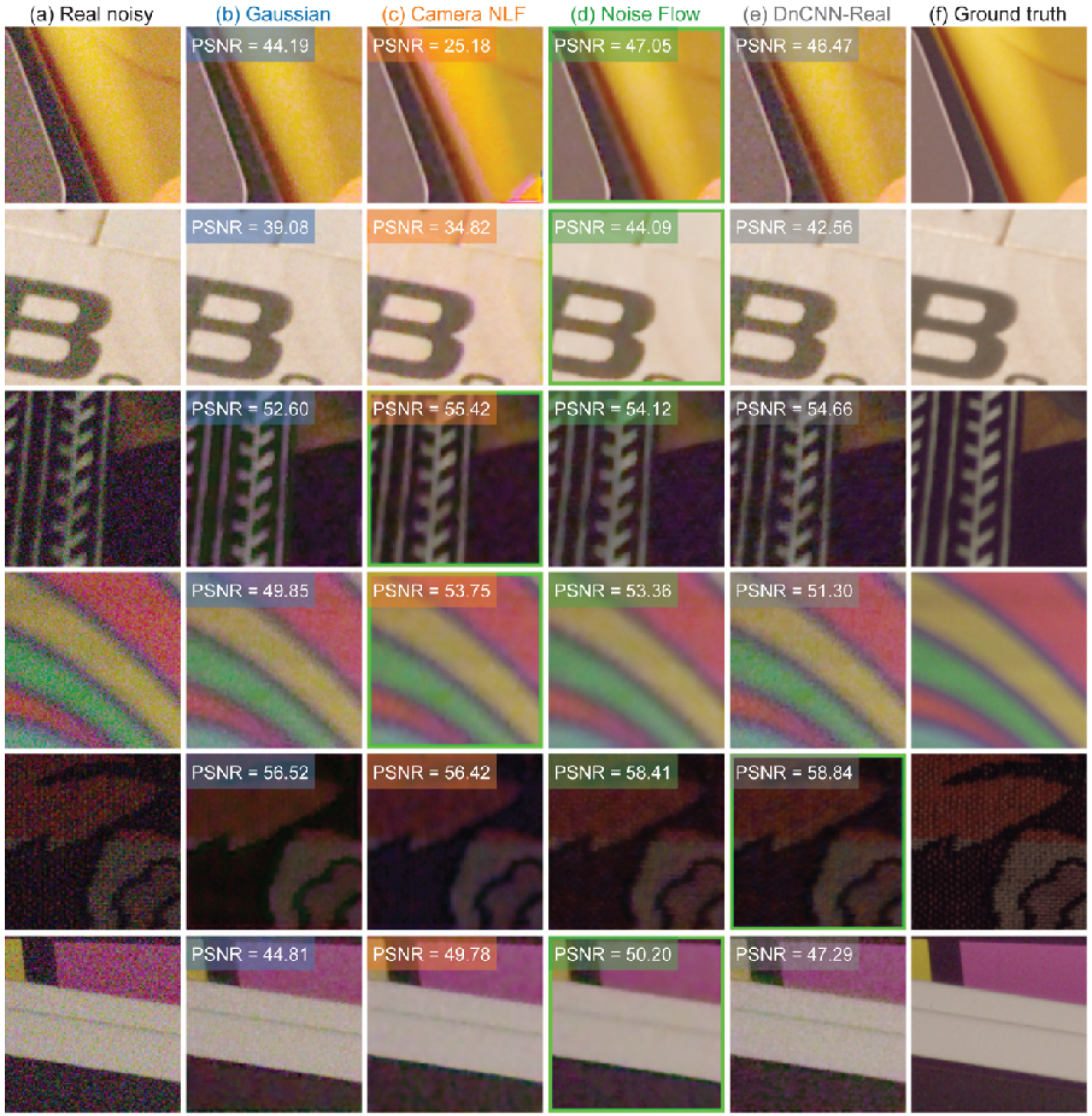
Model Evaluation

	Gaussian	Camera NLF	NoiseFlow
NLL	-2.831	-3.105	-3.521
Marginal KL	0.394	0.052	0.008

Results on SIDD

Denoising with DnCNN

	PSNR (dB)	SSIM
Gaussian	43.63	0.968
Camera NLF	44.99	0.982
NoiseFlow	**48.52**	**0.992**
Real	47.08	0.989



Noise Flow tl;dr

Noise Flow is a realistic and practical model of camera noise in real images

- Domain knowledge to guide construction of model to capture signal dependence, ISO gain and camera specific characteristics
- NFs to learn other aspects which are unknown or difficult to model

Future Directions

- Noise modelling for other sensors and imaging domains
- Other aspects of camera noise (fixed pattern noise, camera specific behaviour, etc)
- Realistic camera noise in other applications

Full details in *Abdelhamed et al ICCV 2019*

Abdelrahman Abdelhamed

Michael S. Brown

Wavelet Flow:
Fast Training of High Resolution Normalizing Flows

Jason J. Yu

Konstantinos G. Derpanis

LASSONDE SCHOOL OF ENGINEERING

YORK UNIVERSITÉ UNIVERSITY

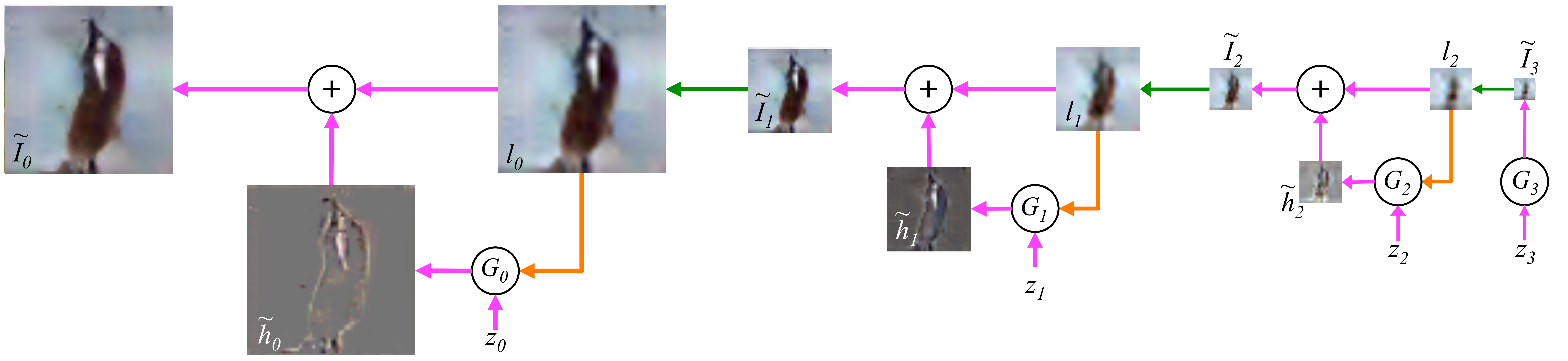
BOREALIS AI

Scale Structure in Generative Models

Existing NF architectures lack explicit notion of *signal* scale

- Models trained at different resolutions are inconsistent
- Training is expensive

GANs and VAEs have exploited image pyramids *[Denton et al 2015; Karras et al 2017]*



Scale Structure in Generative Models

Image pyramids have a long, successful history in computer vision

- **Problem:** Overcomplete

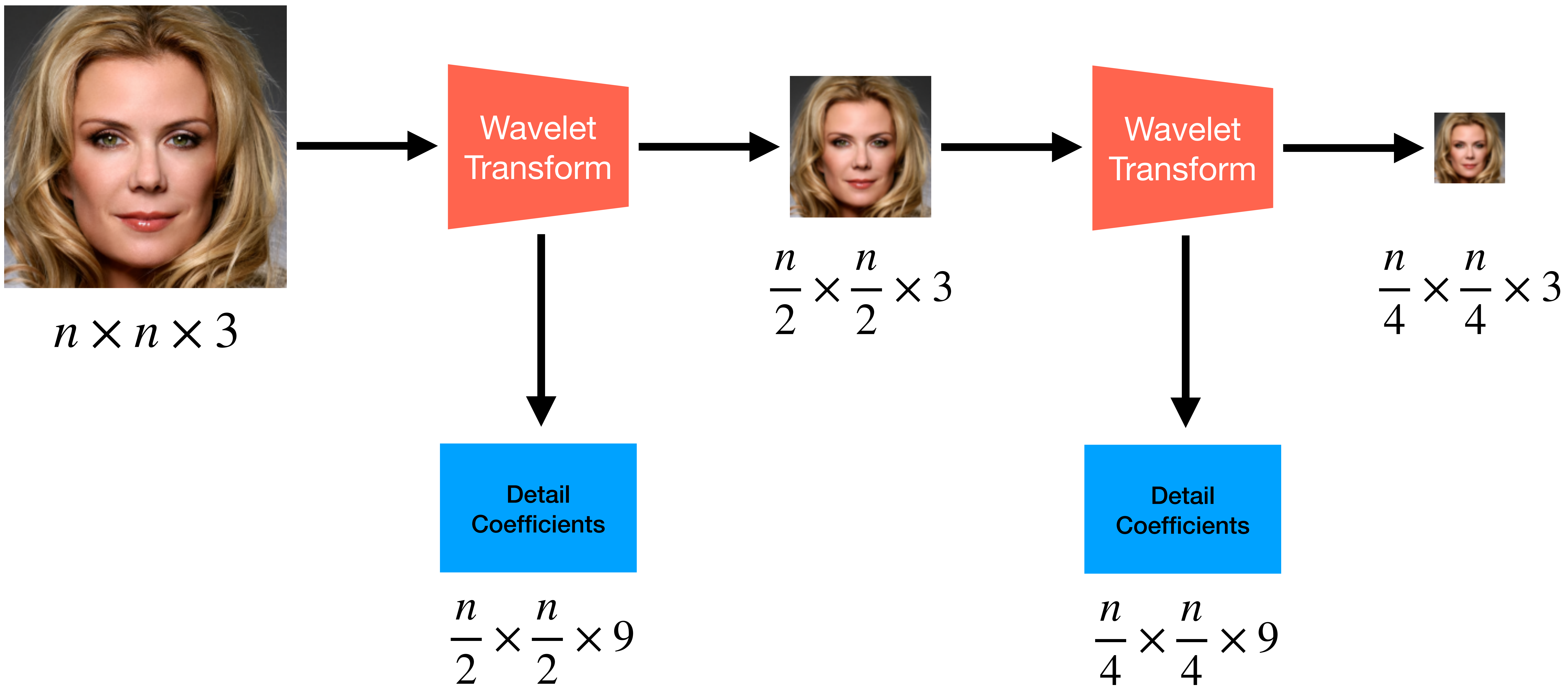
To maintain invertibility and exact density, need to preserve dimensionality

- **Solution:** Wavelets

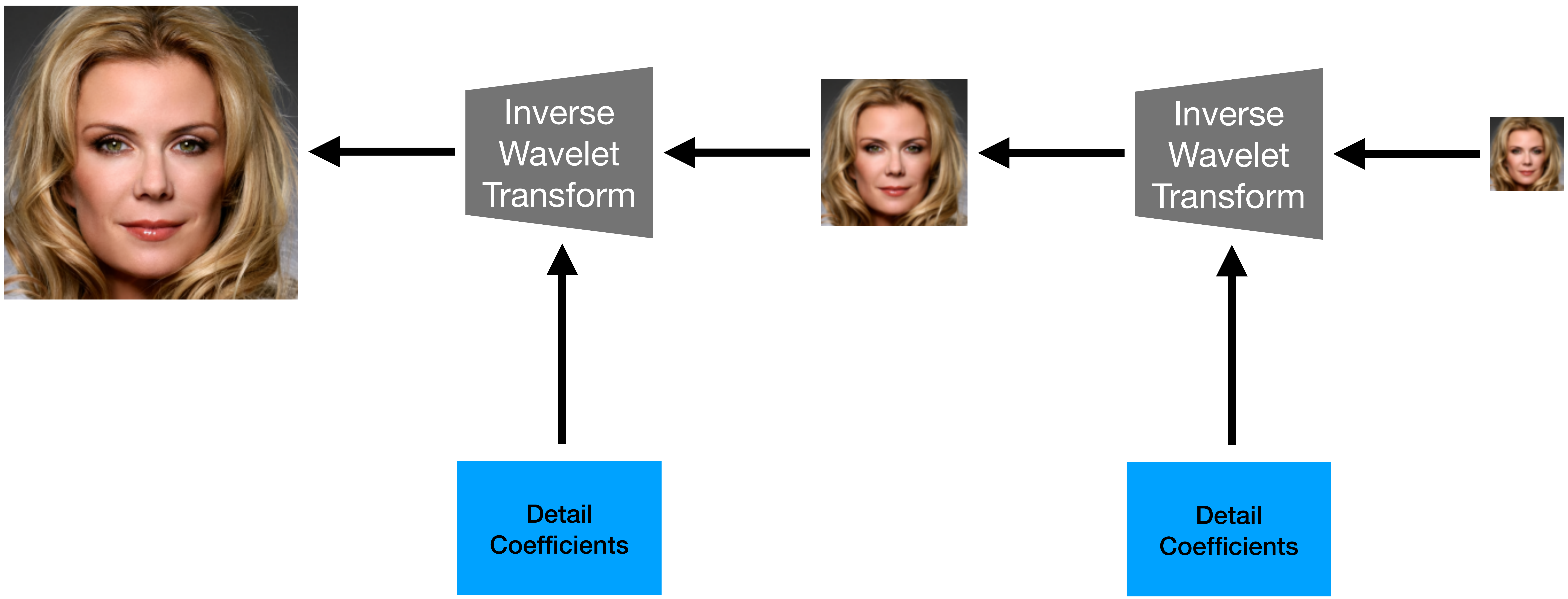
Gaussian pyramid

Laplacian pyramid

Wavelet Transform



Inverse Wavelet Transform



Wavelets

Formally, $\mathbf{I}_0, \mathbf{D}_0, \mathbf{D}_1, \mathbf{D}_2, \ldots, \mathbf{D}_s = h(\mathbf{I})$ where $\mathbf{I} \in \mathbb{R}^{2^{s+1}} \times \mathbb{R}^{2^{s+1}} \times 3$

- $h(\mathbf{I})$ preserves dimensionality => (potentially) invertible
- $h(\mathbf{I})$ is linear => differentiable and constant determinant
- $h(\mathbf{I})$ is orthonormal (for some wavelets) => unit determinant

Can use a wavelet transform as a flow

- In practice used the (Orthonormal) Haar Wavelet

Wavelet Flow

Use change of variables to write

$$
\begin{aligned}
p(\mathbf{I}) &= p(h(\mathbf{I}))\,|\det Dh(I)\,| \\
&= p(\mathbf{I}_0, \mathbf{D}_0, \mathbf{D}_1, \ldots, \mathbf{D}_s)
\end{aligned}
$$

Use product rule of probability to factorize

$$
= p(\mathbf{I}_0)p(\mathbf{D}_0 \,|\, \mathbf{I}_0)p(\mathbf{D}_1 \,|\, \mathbf{D}_0, \mathbf{I}_0)\ldots
$$

Apply inverse wavelet transform $\mathbf{I}_{i+1} = h^{-1}(\mathbf{I}_0, \mathbf{D}_0, \mathbf{D}_1, \ldots, \mathbf{D}_i)$ to get

$$
= p(\mathbf{I}_0)\prod_{i=0}^{s} p(\mathbf{D}_i \,|\, \mathbf{I}_i)
$$

Wavelet Flow

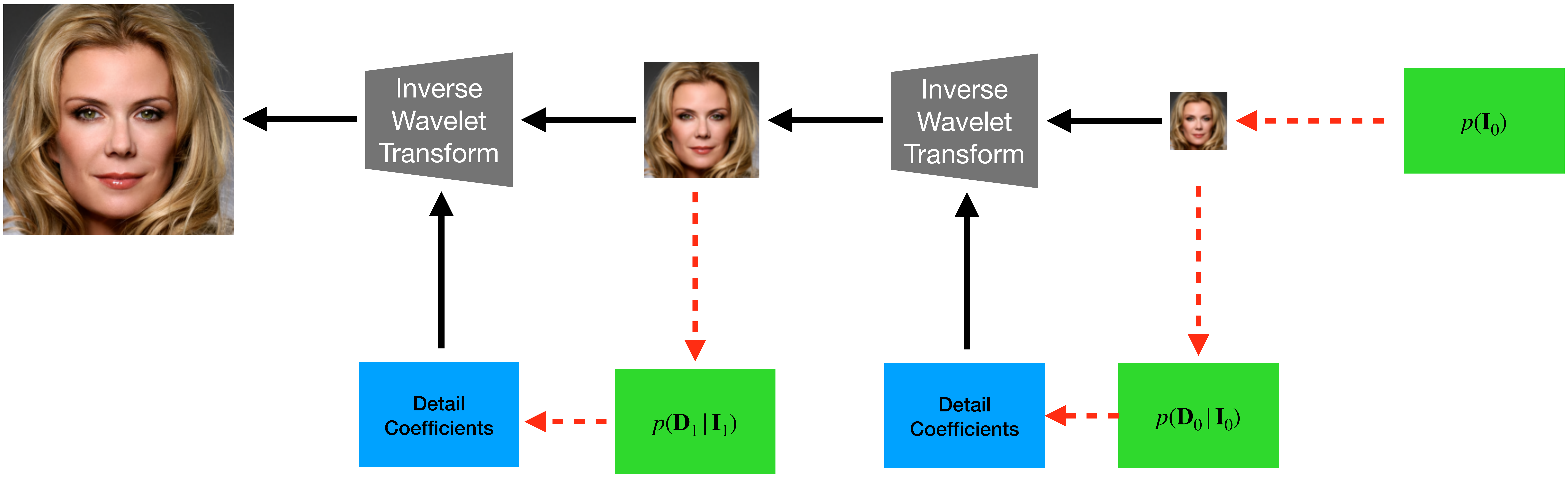
Training can be done with maximum log likelihood but now

$$\log p(\mathbf{I}) = \log p(\mathbf{I}_0) + \sum_{i=0}^{s} \log p(\mathbf{D}_i \,|\, \mathbf{I}_i)$$

The distributions $p(\mathbf{I}_0)$ and $p(\mathbf{D}_i \,|\, \mathbf{I}_i)$ can all be trained independently

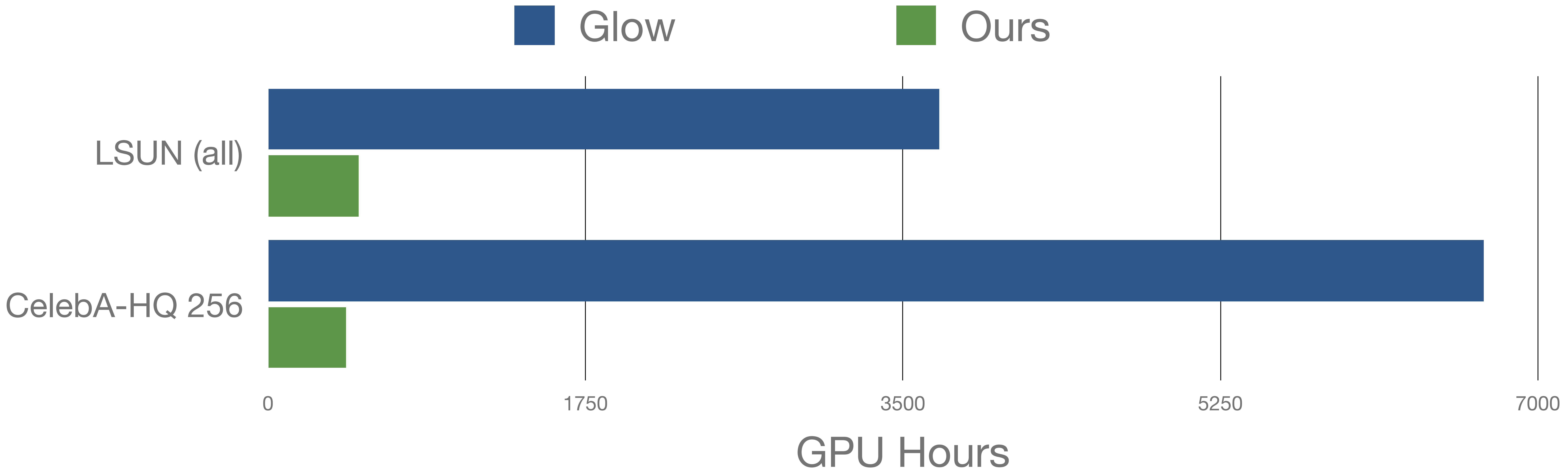
In practice use a Glow-based NF architecture for $p(\mathbf{I}_0)$ and $p(\mathbf{D}_i \,|\, \mathbf{I}_i)$

Generation with Wavelet Flow



Quantitative evaluation

Training time

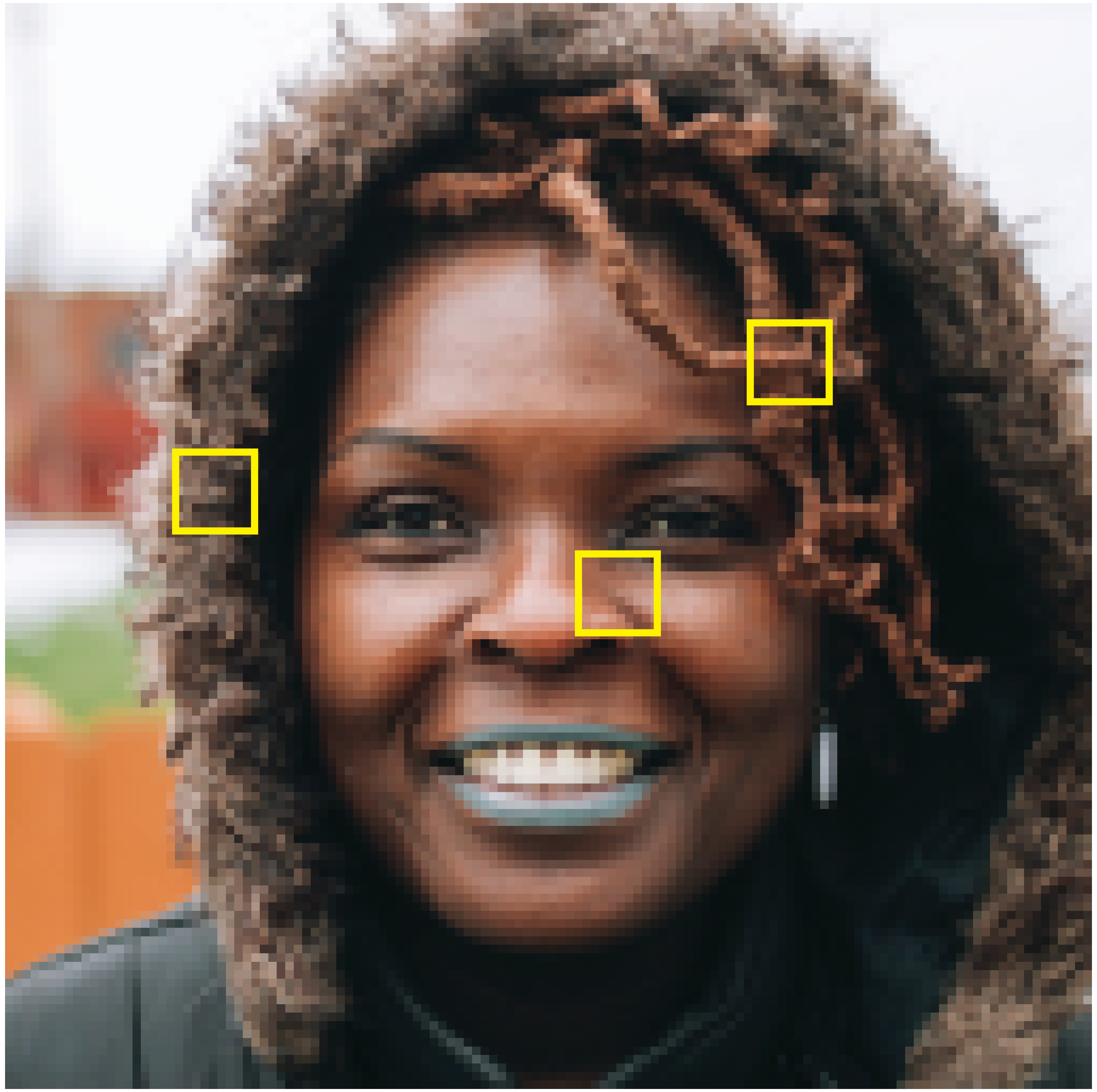


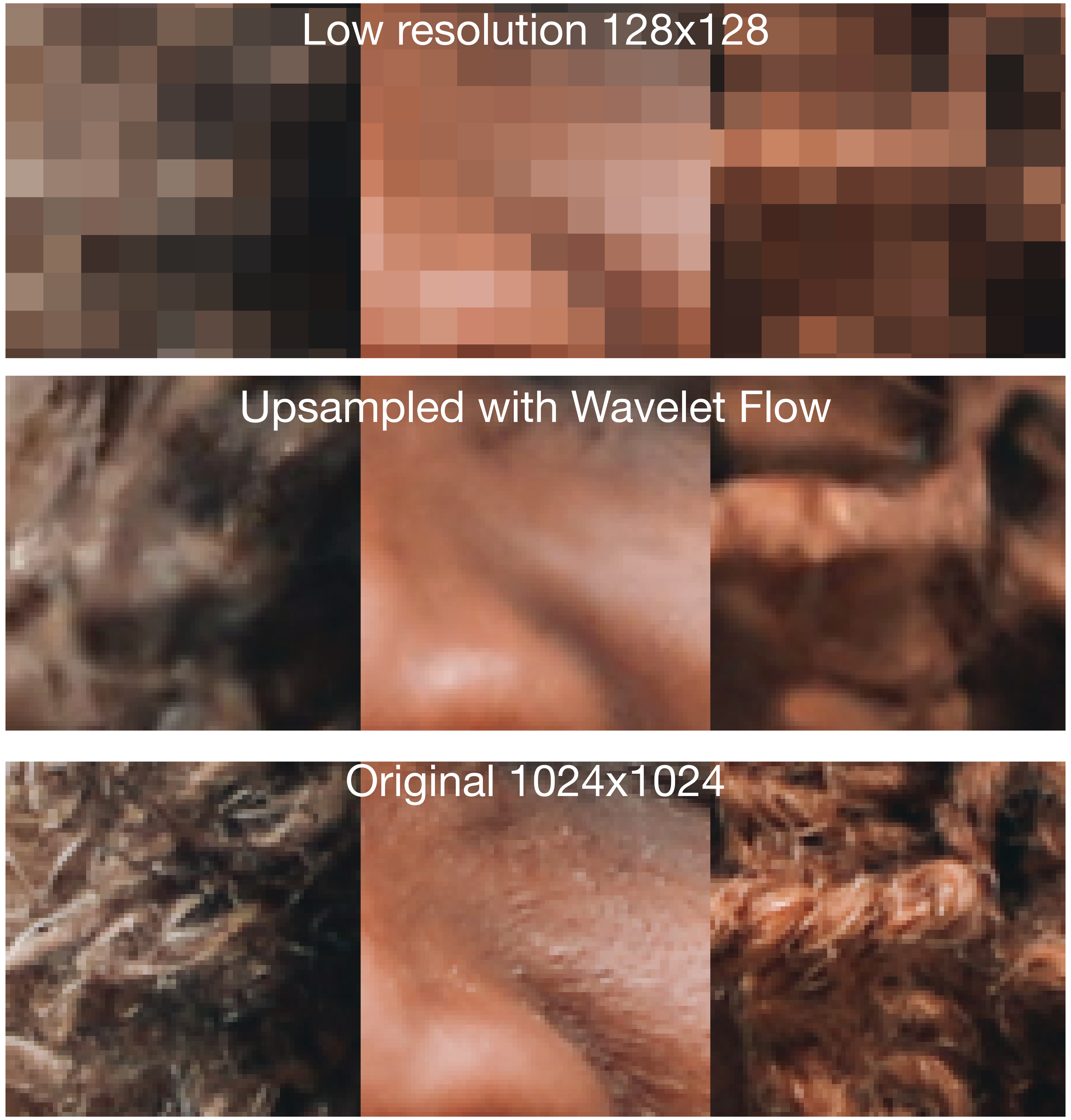
Super-resolution (128x128 Image, 8x upsampled)

Super-resolution (128x128 Image, 8x upsampled w/ Wavelet Flow)

Super-resolution (Original 1024x1024 Image)

Super-resolution Detail Comparison





Wavelet Flow tl;dr

Wavelet Flow:

- Each $p(\mathbf{D}_i \mid \mathbf{I}_i)$ can be simpler and learned independently
- Training can be parallelized for efficient high resolution training (up to 15x faster)
- Every model includes consistent lower resolution models
- Includes super-resolution for free

Limitations and Future Work:

- Perceptual quality is limited, even if quantitatively similar
- Running on other kinds of signals (3D data like MRI/CT/etc)

Full details in Yu et al NeurIPS *2020*

Jason J. Yu

Konstantinos G. Derpanis