HIGH-QUALITY SELF-SUPERVISED DEEP IMAGE DENOISING

Samuli Laine
NVIDIA
slaine@nvidia.com

Tero Karras
NVIDIA
tkarras@nvidia.com

Jaakko Lehtinen
NVIDIA, Aalto University
jlehtinen@nvidia.com

Timo Aila
NVIDIA
taila@nvidia.com

Code and pre-trained models: https://github.com/NVlabs/selfsupervised-denoising

INTRODUCTION

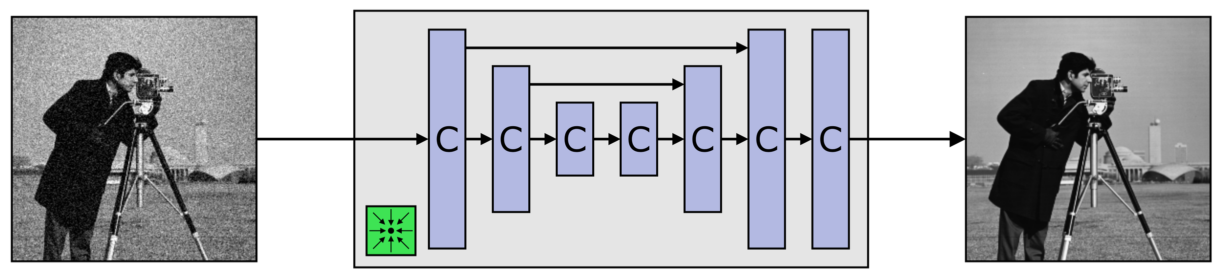
Motivation

- Supervised training requires clean reference data that may not always be available (e.g., cryo-EM, astronomy)
- Goal: Learn to denoise images with no reference data available during training
- How close can we get to standard supervised training that makes use of clean data?

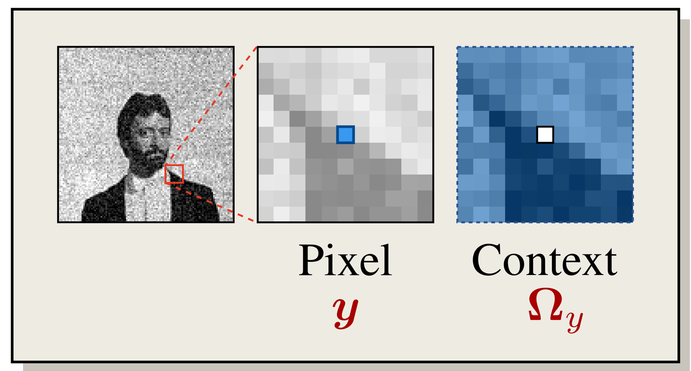
Contributions

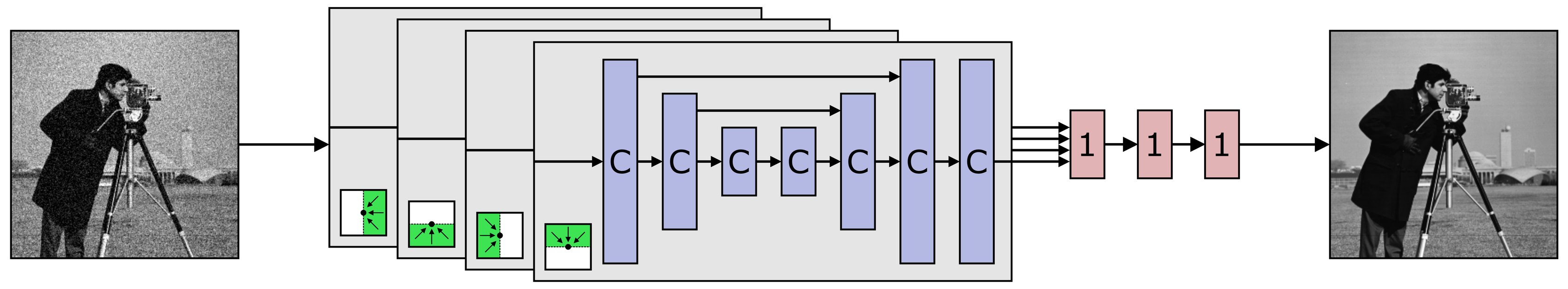
- Convolutional network architecture for constructing blind-spot networks [2] that can be trained efficiently
- Two-step Bayesian inference procedure for combining a learned prior with test-time observations of noisy signal
- Validated on Gaussian, Poisson, and impulse noise
- Denoising quality comparable to supervised training

NETWORK ARCHITECTURE

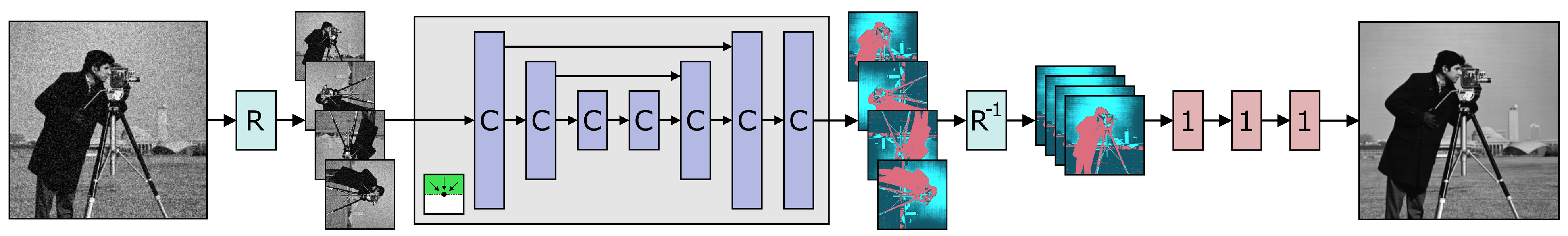


Traditional: U-Net [4]



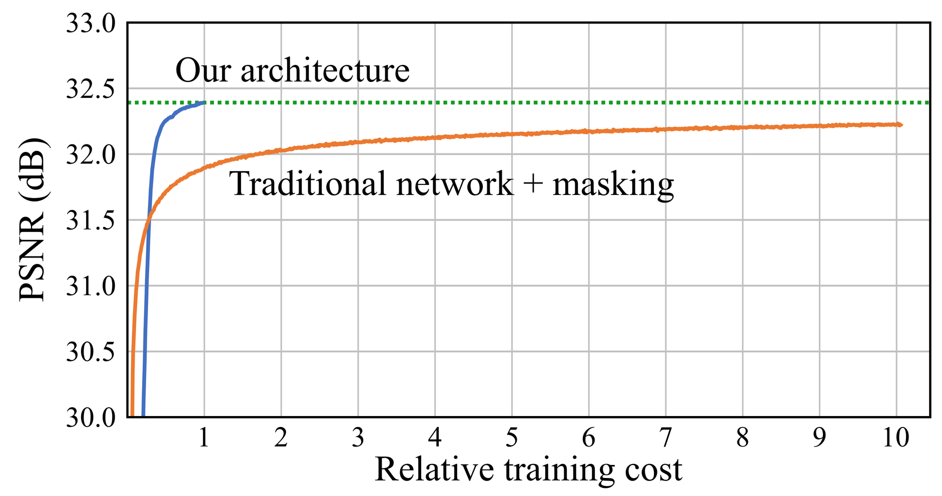


Our convolutional blind-spot: Combine branches with restricted receptive field

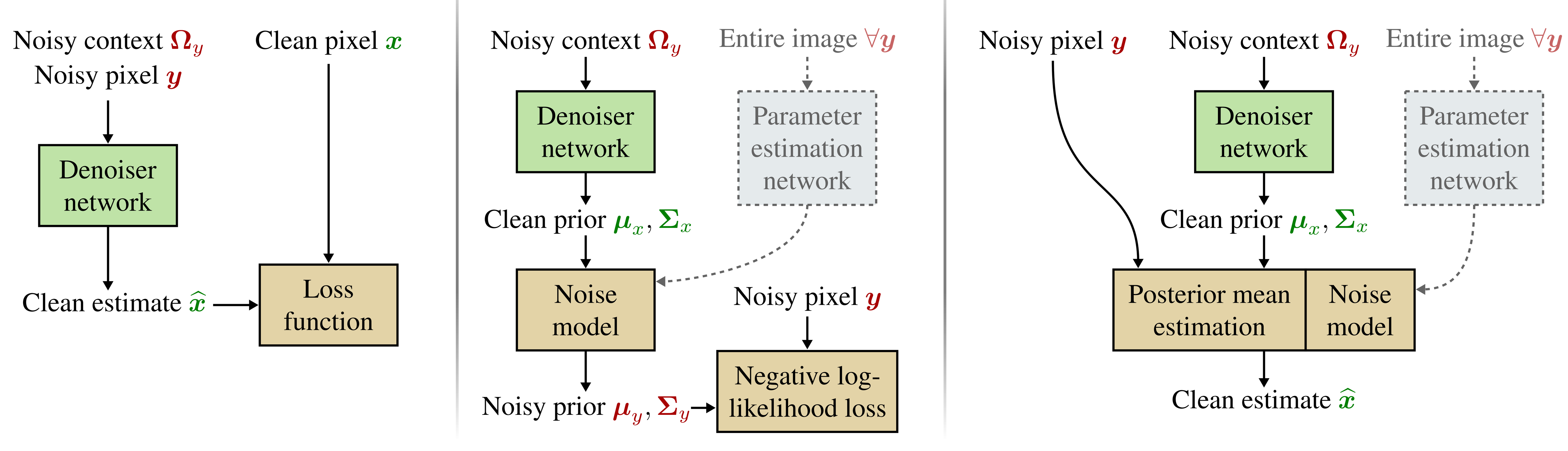


Our optimized variant: Unify branches by rotating data instead

- Our method is based on *blind-spot networks* [2] whose receptive field does not include pixel itself
- We enforce the blind spot via a novel architecture — previously, blind spot was induced during training via *masking* [1, 2, 3]
- We construct branches with half-plane receptive fields, combine to cover all except center pixel



BAYESIAN TRAINING AND INFERENCE



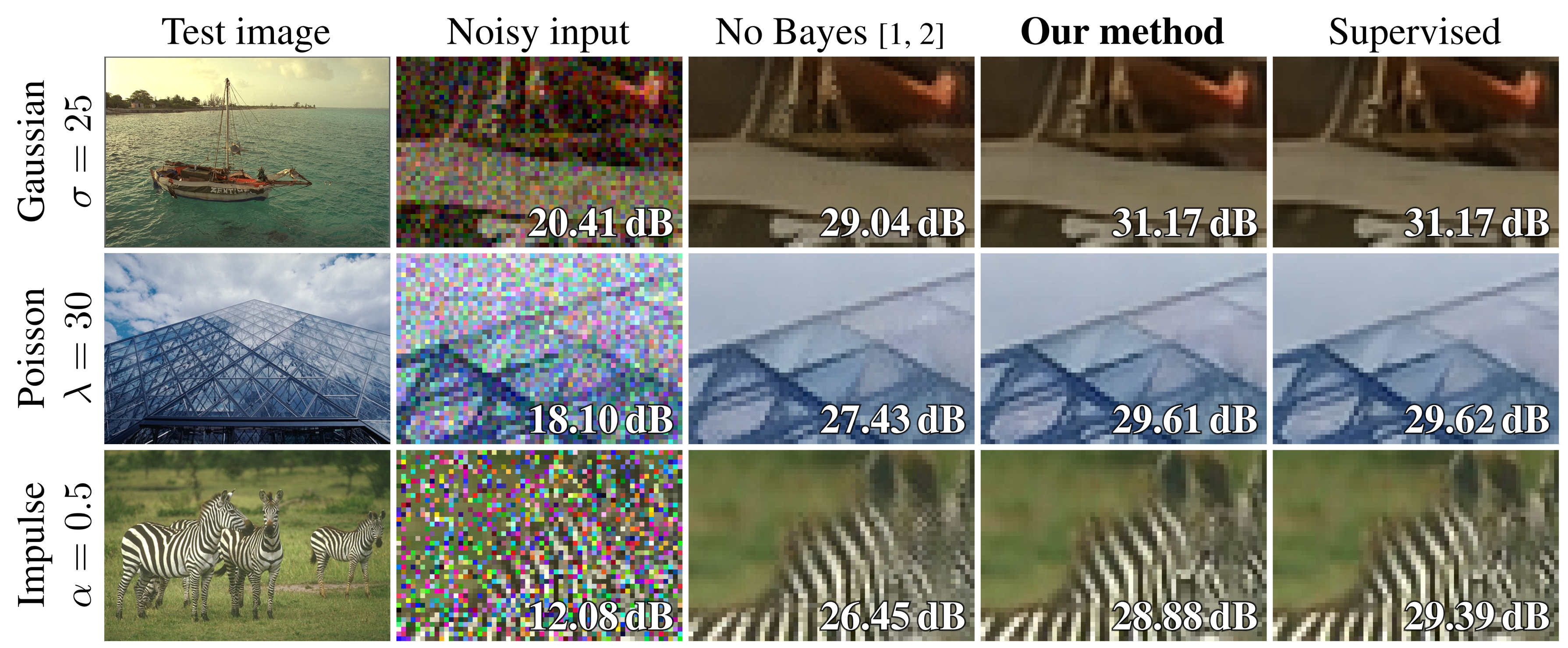
- We learn to predict a Gaussian prior $\mathcal{N}(\boldsymbol{\mu_x}, \boldsymbol{\Sigma_x})$ for clean pixel based on context $\boldsymbol{\Omega_y}$
- Known forward noise model maps this to a prior for noisy pixel $\mathcal{N}(\boldsymbol{\mu_y}, \boldsymbol{\Sigma_y})$ which is fit to observed noisy values $\boldsymbol{y}$ via maximum likelihood loss
- Optional: Auxiliary network is learned to estimate unknown noise model parameters
- For inference, estimated prior $\mathcal{N}(\boldsymbol{\mu_x}, \boldsymbol{\Sigma_x})$ is combined with observed noisy pixel $\boldsymbol{y}$ using posterior mean estimation → minimize expected MSE error → maximize PSNR

Noisy input | Without Bayes [1, 2] | Our method

DENOISING RESULTS

Noise type	Method	Average (dB)
Gaussian $\sigma = 25$	Supervised training	31.60
	Our method*	31.56
	CBM3D	30.96
Gaussian $\sigma \in [5, 50]$	Supervised training	31.71
	Our method	31.59
	CBM3D	31.13
Poisson $\lambda = 30$	Supervised training	30.89
	Our method	30.78
Poisson $\lambda \in [5, 50]$	Supervised training	30.40
	Our method	29.79
Impulse $\alpha = 0.5$	Supervised training	31.98
	Our method	31.57
Impulse $\alpha \in [0, 1]$	Supervised training	30.58
	Our method	30.29

*: Noise parameters assumed unknown in all cases



[1] J. Batson and L. Royer. Noise2Self: Blind denoising by self-supervision. In *Proc. ICML*, 2019.
[2] A. Krull, T.-O. Buchholz, and F. Jug. Noise2Void – Learning denoising from single noisy images. In *Proc. CVPR*, 2019.
[3] A. Krull, T. Vicar, and F. Jug. Probabilistic Noise2Void: Unsupervised content-aware denoising. *CoRR*, abs/1906.00651, 2019.
[4] O. Ronneberger, P. Fischer, and T. Brox. U-Net: Convolutional networks for biomedical image segmentation. *MICCAI*, 9351, 2015.