Normalization Equivariance for Arbitrary Backbones

with Application to Image Denoising

Youssef Saied[1], François Fleuret[1,2]

[1]Department of Computer Science, University of Geneva, Switzerland
[2]Meta FAIR

ICML 2026

Denoisers as Inference Primitives

Inverse problem solvers and diffusion pipelines
call a denoiser $D_\theta(y;\sigma)$ repeatedly

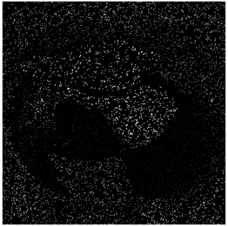
Mismatch: effective noise level outside the training range, or different from the noise level σ passed to the model.

Goal: reliable reuse under noise-level mismatch, without changing the denoiser architecture.

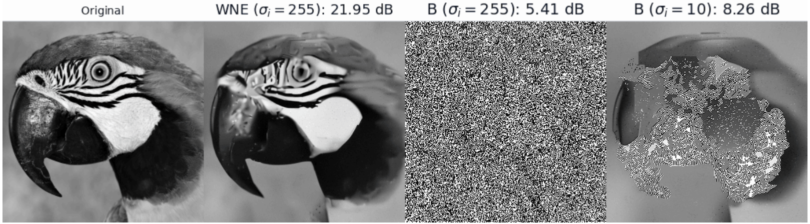
A Denoiser Reused Inside a Sampler

Noise2Noise SwinIR trained at $\sigma = 10$, reused for random inpainting with 10% observed pixels.

observed pixels



sampler trajectory



Set12 PSNR: Standard SwinIR 6 dB   SwinIR-WNE 24 dB

same backbone, training pairs, and sampler hyperparameters

Normalization Equivariance (NE)

$$f(ay + b\mathbf{1}) = af(y) + b\mathbf{1} \qquad a > 0,\ b \in \mathbb{R}$$

For images. A global contrast/brightness change in the input produces the same contrast/brightness change in the output.

Context. Classical denoisers often have this property. Standard CNNs and transformers usually do not.

Existing Methods Are Restrictive

Layer-wise NE enforces equivariance inside every layer:

- constrained convolutions,
- special nonlinearities,
- affine residual connections.

Tradeoff: exact, but invasive.

- rules out standard attention and normalization layers,
- changes the architecture,
- adds runtime overhead.

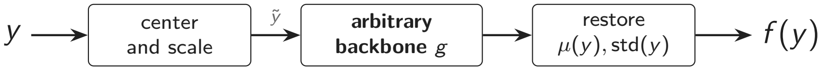
Question: can we get exact NE without changing the denoiser architecture?

Exact NE by Wrapping Any Backbone

Characterization. NE maps have exactly the form below. No NE functions are lost.

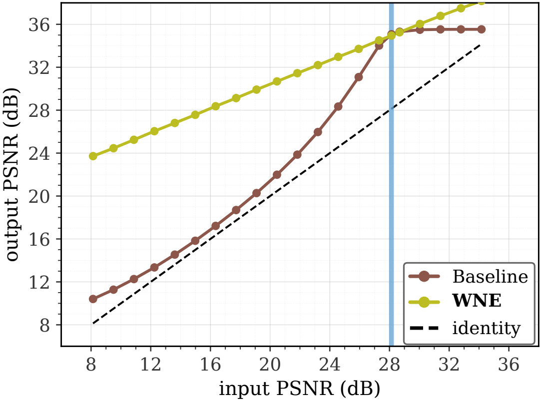
$$f(y) = \mathrm{std}(y)\, g\left(\frac{y - \mu(y)\mathbf{1}}{\mathrm{std}(y)}\right) + \mu(y)\mathbf{1}.$$

Method. Wrap any backbone g for exact equivariance.



WNE: Same Backbone, Stable Mismatch

SwinIR, $\sigma_{\mathrm{train}} = 10$



WNE

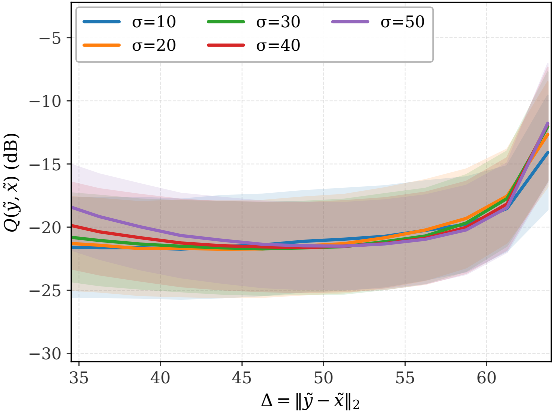
Wrapped Normalization Equivariance

- no additional trainable parameters,
- backbone unchanged,
- attention, LayerNorm, and BatchNorm stay,
- GPU runtime matches baseline; architectural NE is about 1.6× slower.

Why Normalization Helps With Mismatch

Normalized distance: $\Delta := \|\tilde{y} - \tilde{x}\|_2$

Q: normalized denoising quality, higher is better



Normalized coordinates.

WNE makes the backbone process normalized images.

Main pattern.

At the same Δ, curves for different σ mostly line up.

Takeaway.

Noise-level mismatch becomes a more stable normalized regression problem.