High Perceptual Quality Image Denoising with a Posterior Sampling CGAN

Guy Ohayon† Theo Adrai† Gregory Vaksman† Michael Elad‡ Peyman Milanfar‡

†Technion, Israel Institute of Technology ‡Google Research

Contributions

- A stochastic image denoising scheme producing visually pleasing images while ensuring high PSNR performance
- A revision of CGAN training framework which alleviates mode collapse
- A simple way to traverse the perception-distortion tradeoff at inference time

Background

Problem Statement: Given a noisy image $\mathbf{y}$ of a clean source $\mathbf{x}$, obtain $\hat{\mathbf{x}}$, a natural looking image "close" to $\mathbf{x}$

Solution: Sample from the posterior distribution $\hat{\mathbf{x}} \sim \mathbb{P}_{\mathbf{x}|\mathbf{y}}$, which is known to produce perfect perceptual quality images with MSE performance twice that of the MMSE estimator [1]

Method

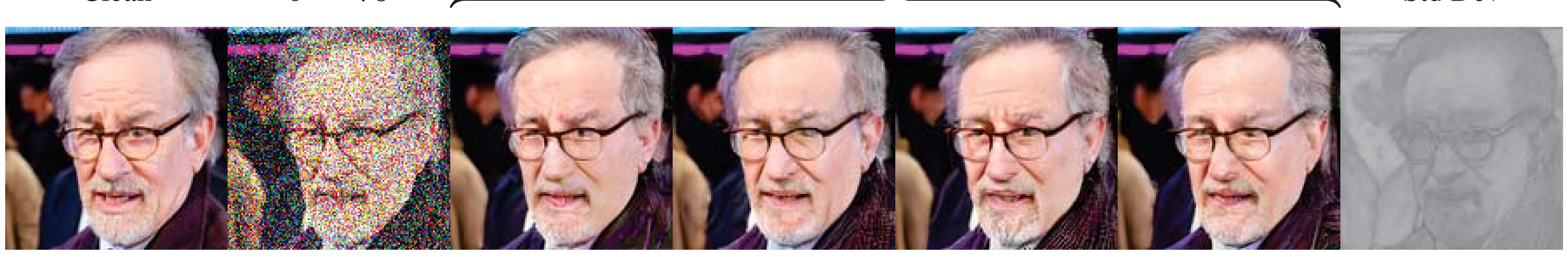
Loss: We sample from the posterior by optimizing the CGAN objective [3] $\mathcal{L}_{CGAN}$. We introduce an additional penalty term which maintains the global optimum of $\mathcal{L}_{CGAN}$ but significantly eases training and avoids mode collapse:

$$\theta^* = \arg\min_{\theta} \mathcal{L}_{CGAN} + \lambda \mathbb{E}_{\mathbf{x},\mathbf{y}} \left[\|\mathbf{x} - \mathbb{E}_{\mathbf{z}} \left[G_\theta(\mathbf{z}, \mathbf{y}) | \mathbf{y} \right] \|_2^2 \right]$$

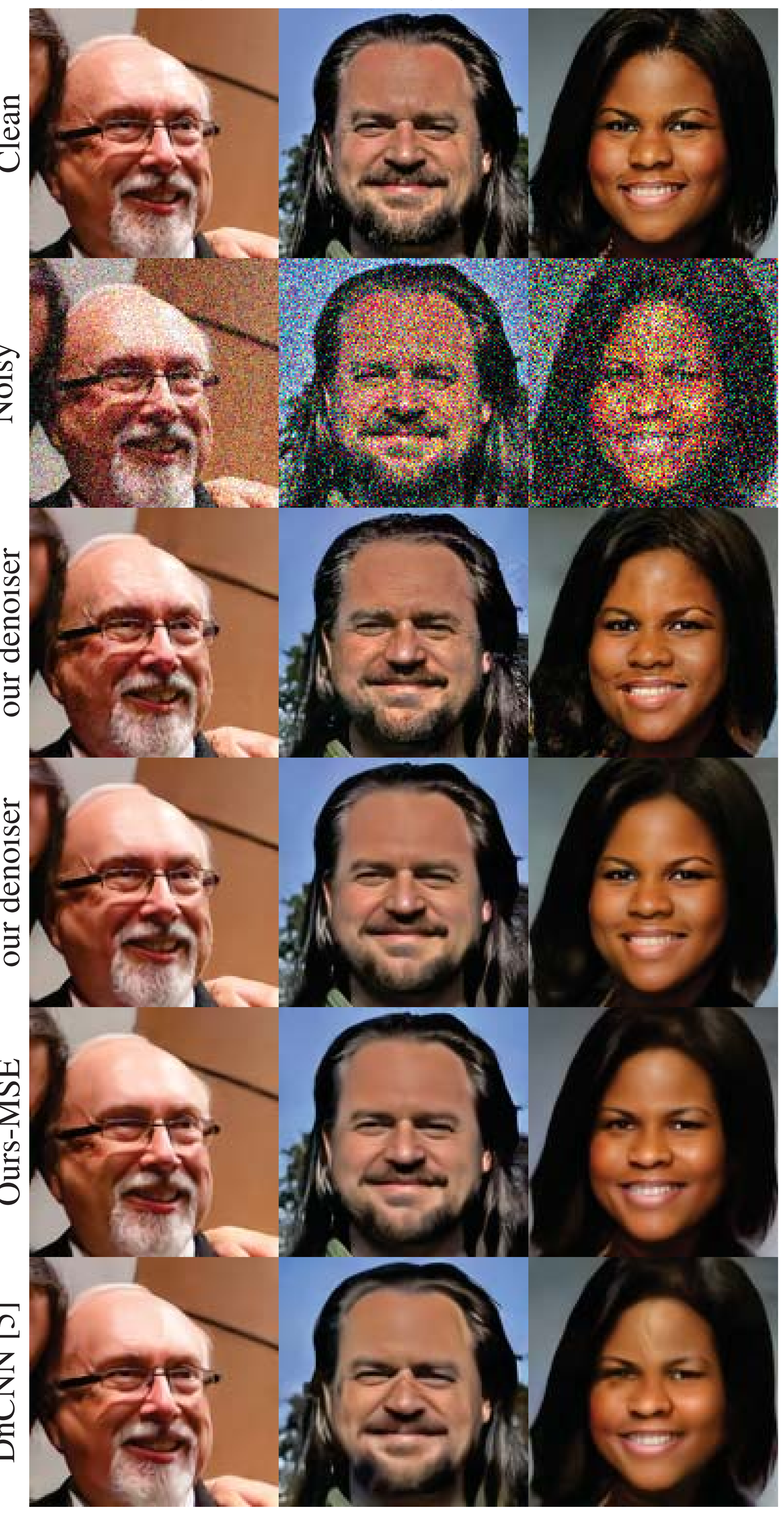
Architecture: Inspired by StyleGAN2 [2] and UNet [4], our denoiser is an encoder-decoder deep neural network:

- The encoder gradually down-scales the input image to extract both global and local details. The features of each down-scaling stage are passed to the corresponding stage of the decoder
- The decoder transforms the features of each stage into a "residual" RGB image, which is added to the up-sampled version of the RGB image from the previous stage
- Stochastic noise is injected at each stage of the decoder to generate scale-specific details in the denoised image

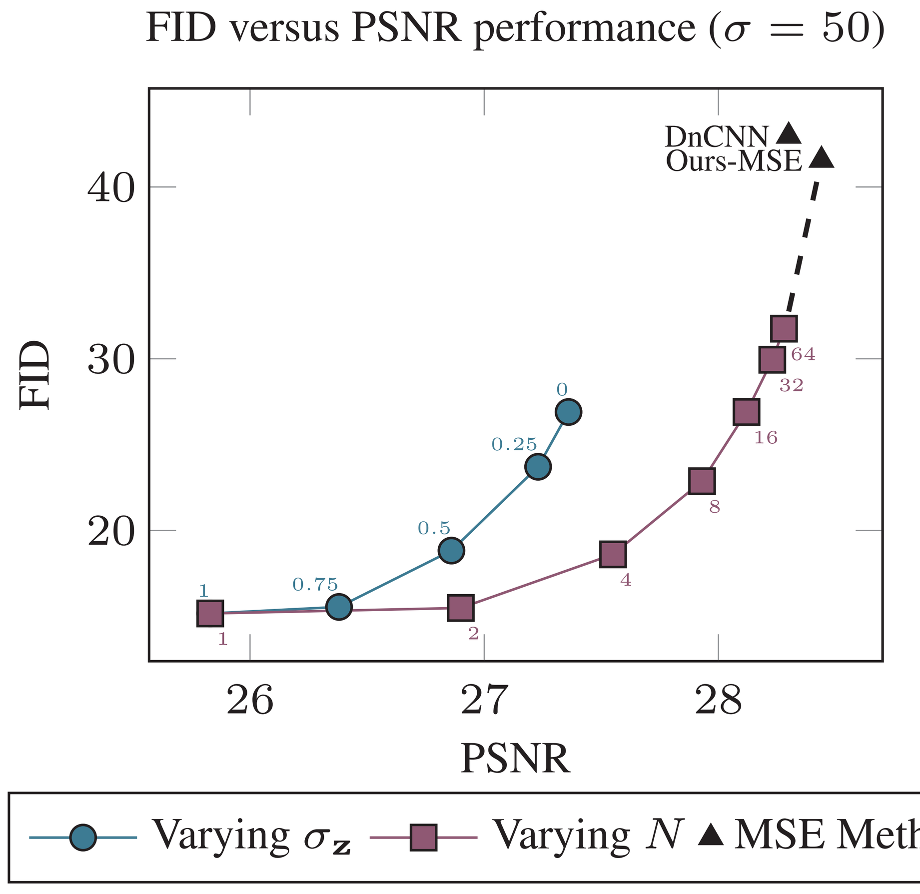
Experiments: stochastic variation of our denoiser



Experiments: visual comparison



Traversing the percetion-distortion tradeoff



We traverse the perception distortion tradeoff by varying at inference time:

- $\sigma_{\mathbf{z}}$, the standard deviation of the noise injected to our stochastic denoiser
- N, the number of instances that are being averaged to produce the resulting image

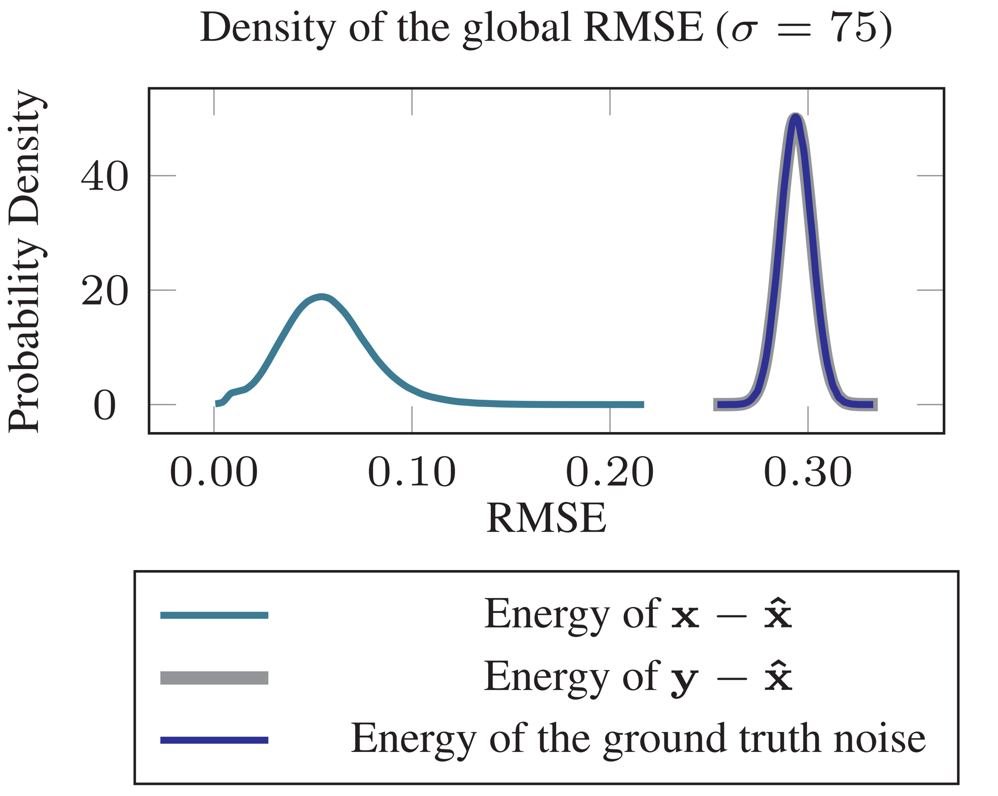
PSNR and FID comparison

σ	One instance		64 instances mean		Ours-MSE		DnCNN	
	PSNR	FID	PSNR	FID	PSNR	FID	PSNR	FID
25	29.19	**12.66** ± 0.07	31.46	27.48	**31.83**	31.48	31.77	36.80
50	25.83	**15.18** ± 0.15	28.28	31.81	**28.44**	41.56	28.30	42.97
75	24.09	**15.78** ± 0.13	26.57	34.64	**26.81**	46.31	26.46	47.69

Did we really obtain a denoiser?

We confirm that our denoiser does not generate improper details by assessing the following properties both globally and locally:

- The remainder noise $\mathbf{y} - \hat{\mathbf{x}}$ should be normally distributed
- The "energy" (L_2 norm) of the residual $\mathbf{x} - \hat{\mathbf{x}}$ should be much lower than that of the ground truth noise



References

[1] Yochai Blau and Tomer Michaeli. The perception-distortion tradeoff. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, June 2018.

[2] Tero Karras et al. Analyzing and improving the image quality of stylegan. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, June 2020.

[3] Mehdi Mirza and Simon Osindero. Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784*, 2014.

[4] O. Ronneberger, P.Fischer, and T. Brox. U-net: convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention*, volume 9351 of *LNCS*, pages 234–241. Springer, 2015.

[5] Kai Zhang et al. Beyond a gaussian denoiser: residual learning of deep cnn for image denoising. *IEEE Transactions on Image Processing*, 26(7):3142–3155, 2017.

TECHNION | The Henry and Marilyn Taub Faculty of Computer Science