Stochastic Image Denoising by Sampling from the Posterior Distribution (Supplementary Material)

Bahjat Kawar Gregory Vaksman Michael Elad
Computer Science Department, The Technion - Israel Institute of Technology
{bahjat.kawar,grishav,elad}@cs.technion.ac.il

1. Experimental Details

Table 1 and Table 2 present the hyperparameters used in our experiments. The hyperparameter σ_L is set to 0.01 in all experiments.

Dataset	ϵ	$\mathbf{T}$	$\frac{\sigma_{\mathbf{i+1}}}{\sigma_{\mathbf{i}}}$
CelebA [2]	$3.3e-6$	5	0.982
LSUN [3]	$1.8e-6$	3	0.991
FFHQ [1]	$1.8e-6$	3	0.995

Table 1. Hyperparameters for our denoising experiments. The last column is the geometric progression rate for $\{\sigma_i\}_{i=0}^{L+1}$.

Dataset	ϵ	$\mathbf{T}$	$\mathbf{L+K}$	$\sigma_{-\mathbf{K}}$	$\frac{\sigma_{\mathbf{i+1}}}{\sigma_{\mathbf{i}}}$
CelebA [2]	$3.3e-6$	5	500	90	0.982
LSUN [3]	$1.8e-6$	3	1086	190	0.991
FFHQ [1]	$1.8e-6$	5	2311	348	0.995

Table 2. Hyperparameters for our inpainting experiments.

2. Perceptual Quality Evaluation

We evaluate the perceptual quality of the results using LPIPS [4] (version 0.1), a distance metric shown to perform better as a perceptual metric than PSNR or SSIM. We evaluated 64 sets of images, each set containing an original CelebA image, a denoised one using an MMSE denoiser, and our algorithm's output, utilizing the same MMSE denoiser. Our model performs significantly better than the MMSE denoiser, achieving around 16% lower (and thus, better) LPIPS perceptual distance from the original image, as can be seen in Table 3.

Dataset	σ_0	Ours	MMSE	Improvement
CelebA [2]	0.203	0.027	0.032	15.6%
	0.406	0.049	0.059	16.9%
	0.607	0.067	0.081	17.3%

Table 3. Average LPIPS distance using 64 CelebA images. The last column shows the percentage of improvement that our algorithms presents.

3. Inpainting Approximate Derivation

The following approximation is used in the derivation of Equation 12, referring to the inpainting algorithm:

$$\nabla_{\tilde{x}^M} \log p\left(\tilde{x}^R|\tilde{x}^M\right) = \nabla_{\tilde{x}^M} \log p\left(\tilde{x}^R|y^M\right) \approx 0. \quad (1)$$

We support this approximation by making the following observation:

$$\begin{aligned}
&\nabla_{\tilde{x}^M} \log p\left(\tilde{x}^R|\tilde{x}^M\right) = \nabla_{\tilde{x}^M} \log \frac{p\left(\tilde{x}^R, \tilde{x}^M\right)}{p\left(\tilde{x}^M\right)} \\
&= \nabla_{\tilde{x}^M} \log p\left(\tilde{x}\right) - \nabla_{\tilde{x}^M} \log p\left(\tilde{x}^M\right) \quad (2) \\
&= \left(\nabla_{\tilde{x}} \log p\left(\tilde{x}\right)\right)^M - \nabla_{\tilde{x}^M} \log p\left(\tilde{x}^M\right) \\
&= \mathbb{E}\left[x|\tilde{x}\right]^M - \mathbb{E}\left[x^M|\tilde{x}^M\right] = \mathbb{E}\left[x^M|\tilde{x}\right] - \mathbb{E}\left[x^M|\tilde{x}^M\right].
\end{aligned}$$

The second to last equality holds due to Equation 8 in the paper. We obtained a difference between two estimators of x^M, both of which depend on information from a noisy version of it, $\tilde{x}^M$, but the second one contains additional information about $\tilde{x}^R$. While this information may change the estimation, we assume its impact to be negligible, resulting in this difference to be approximately zero, thus obtaining Equation 1.

4. Additional Results

In the following, we present additional qualitative results not shown in the paper. They are best viewed digitally, and zoomed in. Standard deviations, p-values, and correlation coefficients were rounded to 2 decimal places.

Figures 1-20 present additional material referring to our denoising scheme. Figure 1 presents both MMSE and our denoising results for several noise levels, demonstrating the tendency of our method to produce far sharper images. Figures 2-20 show further denoising results for images taken from FFHQ and LSUN with varying noise levels, this time emphasizing the possible diversity of the outcomes of our stochastic denoiser, and their validity as denoising results.

Figures 21-24 introduce additional material referring to our inpainting scheme. Figures 21-23 showcase additional

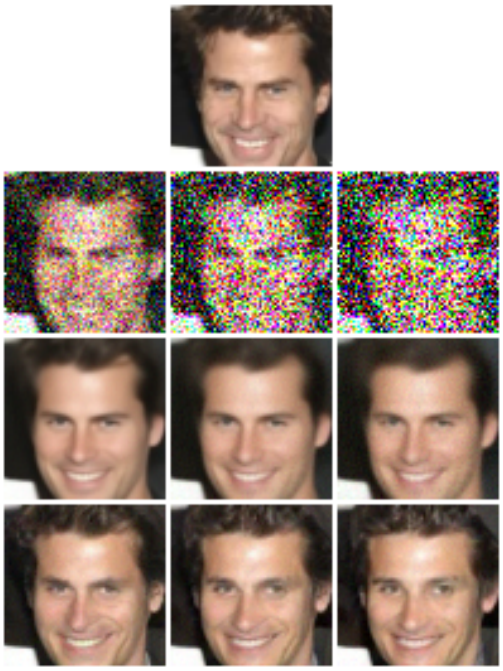

Figure 1. From top to bottom: original CelebA image, noisy versions (σ_0 from left to right: 0.203, 0.406, 0.607), MMSE denoiser outputs, and instances of our denoising algorithm's output.

inpainting results, highlighting the sharpness of the results and the posterior distribution score function's ability to synthesize features consistent with the surroundings of the missing pixels. Figure 24 shows intermediate results obtained along our inpainting algorithm for different images, gradually converging towards the final resulting images.

References

[1] Tero Karras, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. Analyzing and improving the image quality of stylegan. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8110–8119, 2020.

[2] Ziwei Liu, Ping Luo, Xiaogang Wang, and Xiaoou Tang. Deep learning face attributes in the wild. In *Proceedings of the IEEE international conference on computer vision*, pages 3730–3738, 2015.

[3] Fisher Yu, Ari Seff, Yinda Zhang, Shuran Song, Thomas Funkhouser, and Jianxiong Xiao. Lsun: Construction of a large-scale image dataset using deep learning with humans in the loop. *arXiv preprint arXiv:1506.03365*, 2015.

[4] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *CVPR*, 2018.

Figure 2. From top to bottom: original FFHQ images, noisy versions with $\sigma_0 = 0.2$, MMSE denoiser outputs, and two instances of our denoising algorithm's output.

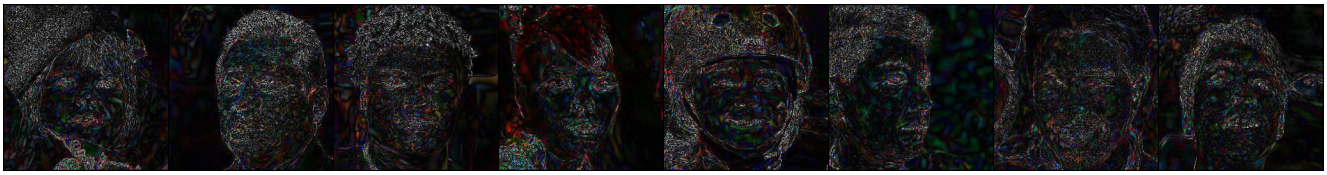

Figure 3. The absolute value of the difference (scaled by 4) between two instances of our denoising algorithm's output on FFHQ images with $\sigma_0 = 0.2$.

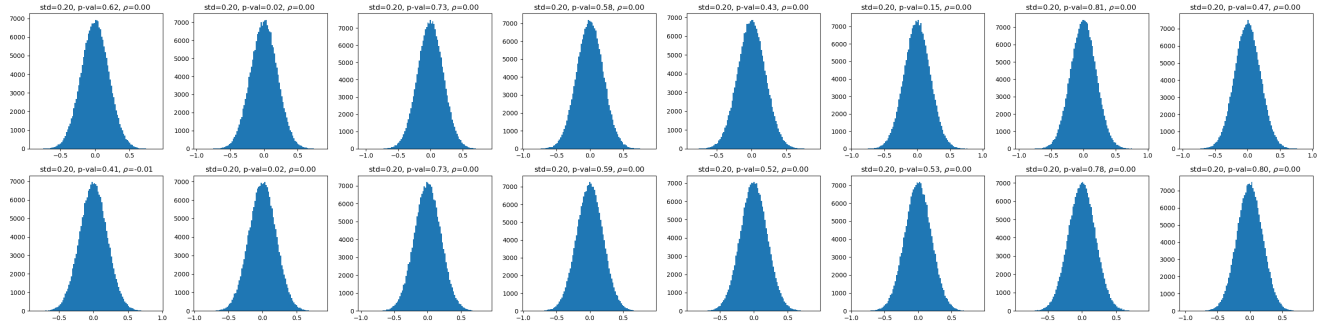

Figure 4. Residual histograms, standard deviations, normality p-values, and the Pearson's correlation coefficients (in the direction with the maximum absolute value) for our denoising algorithm's output on FFHQ images with $\sigma_0 = 0.2$.

Figure 5. From top to bottom: original FFHQ images, noisy versions with $\sigma_0 = 0.4$, MMSE denoiser outputs, and two instances of our denoising algorithm's output.

Figure 6. The absolute value of the difference (scaled by 4) between two instances of our denoising algorithm's output on FFHQ images with $\sigma_0 = 0.4$.

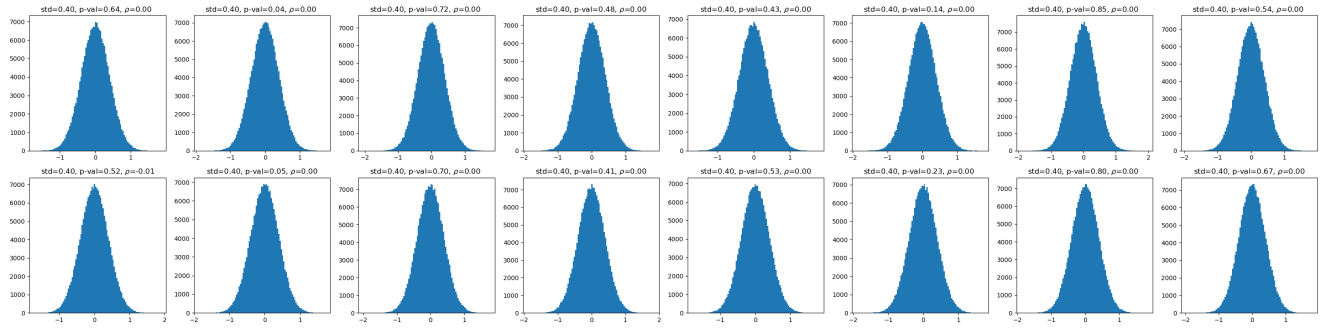

Figure 7. Residual histograms, standard deviations, normality p-values, and the Pearson's correlation coefficients (in the direction with the maximum absolute value) for our denoising algorithm's output on FFHQ images with $\sigma_0 = 0.4$.

Figure 8. From top to bottom: original FFHQ images, noisy versions with $\sigma_0 = 0.602$, MMSE denoiser outputs, and two instances of our denoising algorithm's output.

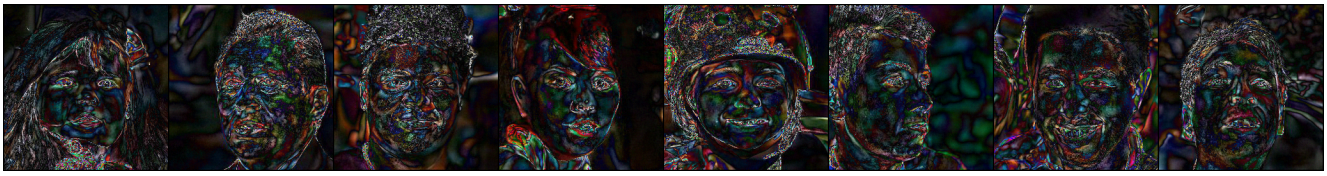

Figure 9. The absolute value of the difference (scaled by 4) between two instances of our denoising algorithm's output on FFHQ images with $\sigma_0 = 0.602$.

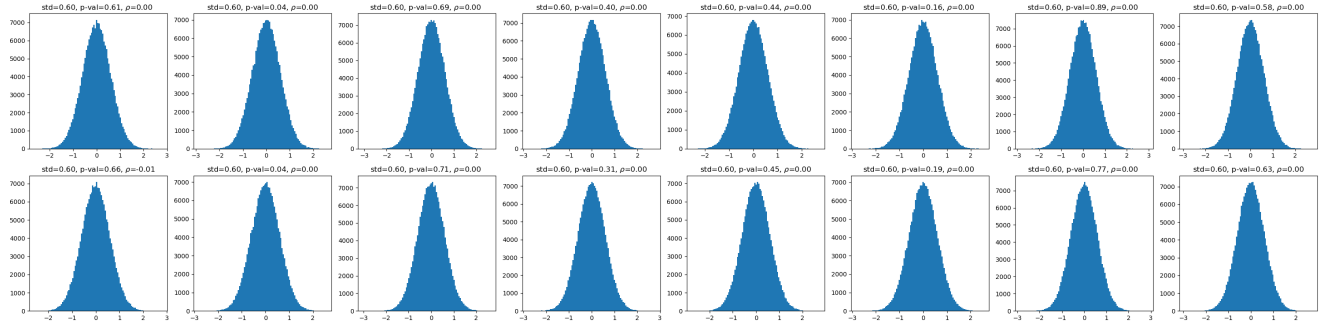

Figure 10. Residual histograms, standard deviations, normality p-values, and the Pearson's correlation coefficients (in the direction with the maximum absolute value) for our denoising algorithm's output on FFHQ images with $\sigma_0 = 0.602$.

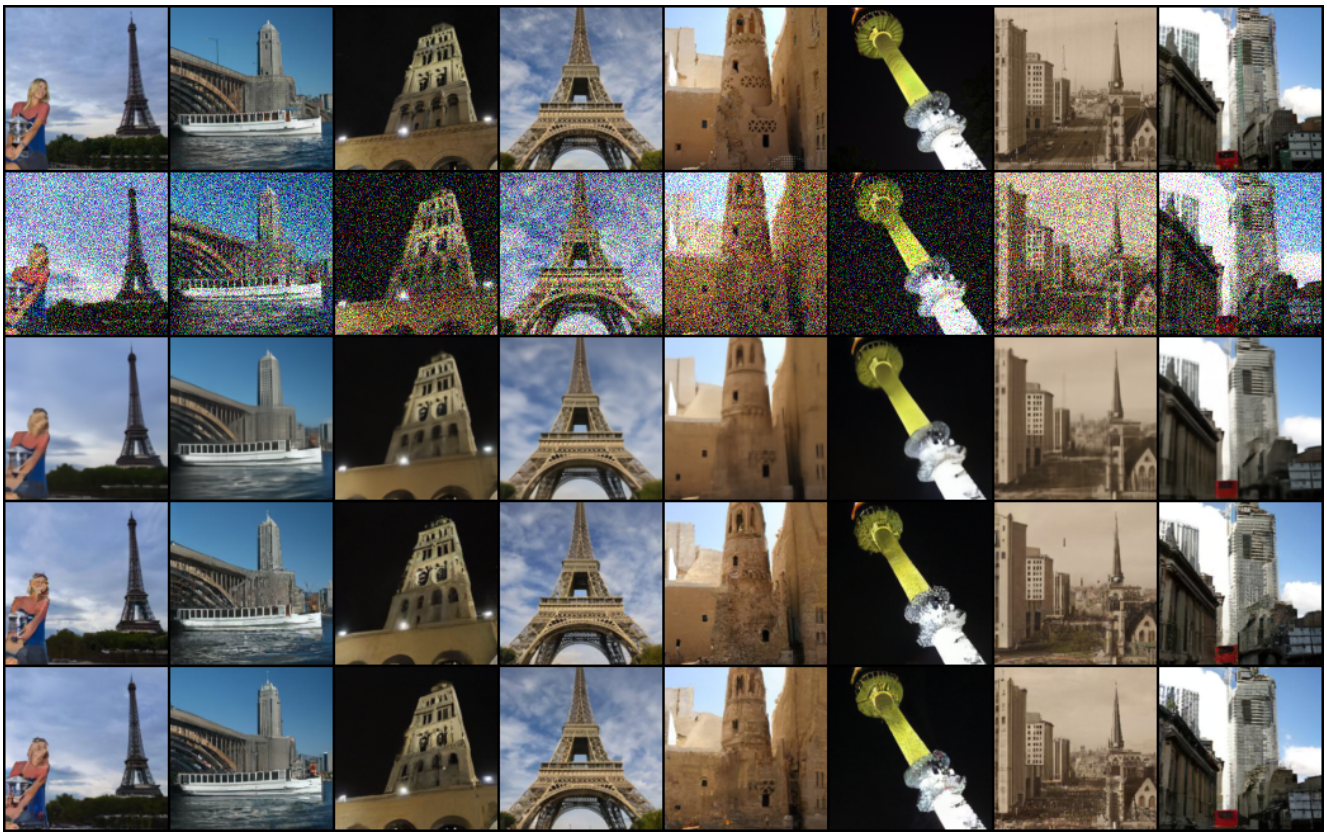

Figure 11. From top to bottom: original LSUN-tower images, noisy versions with $\sigma_0 = 0.198$, MMSE denoiser outputs, and two instances of our denoising algorithm's output.

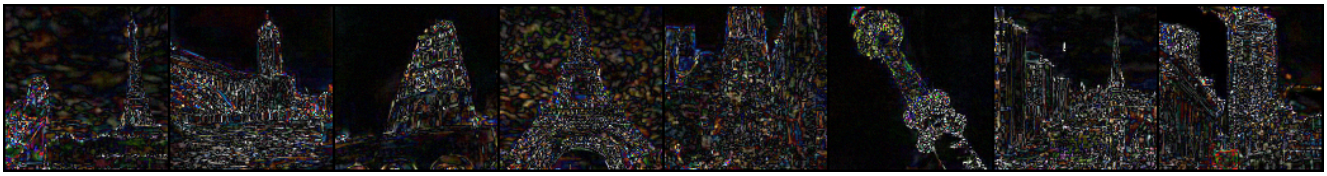

Figure 12. The absolute value of the difference (scaled by 4) between two instances of our denoising algorithm's output on LSUN-tower images with $\sigma_0 = 0.198$.

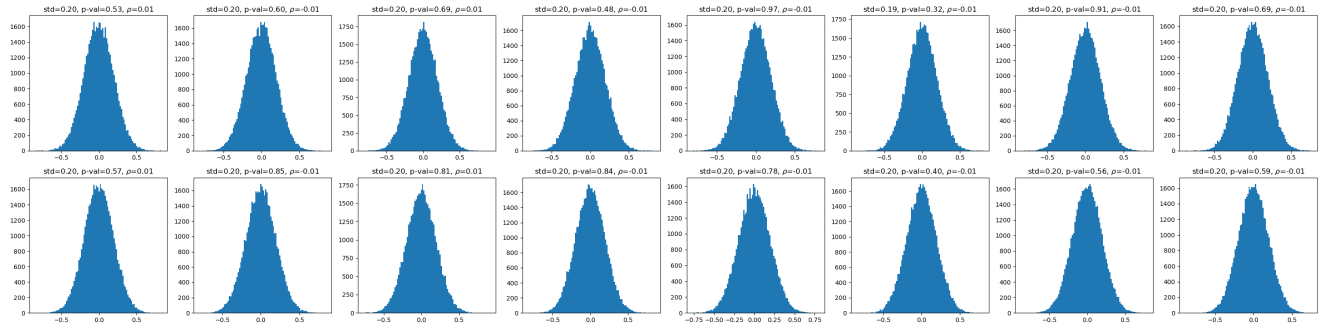

Figure 13. Residual histograms, standard deviations, normality p-values, and the Pearson's correlation coefficients (in the direction with the maximum absolute value) for our denoising algorithm's output on LSUN-tower images with $\sigma_0 = 0.198$.

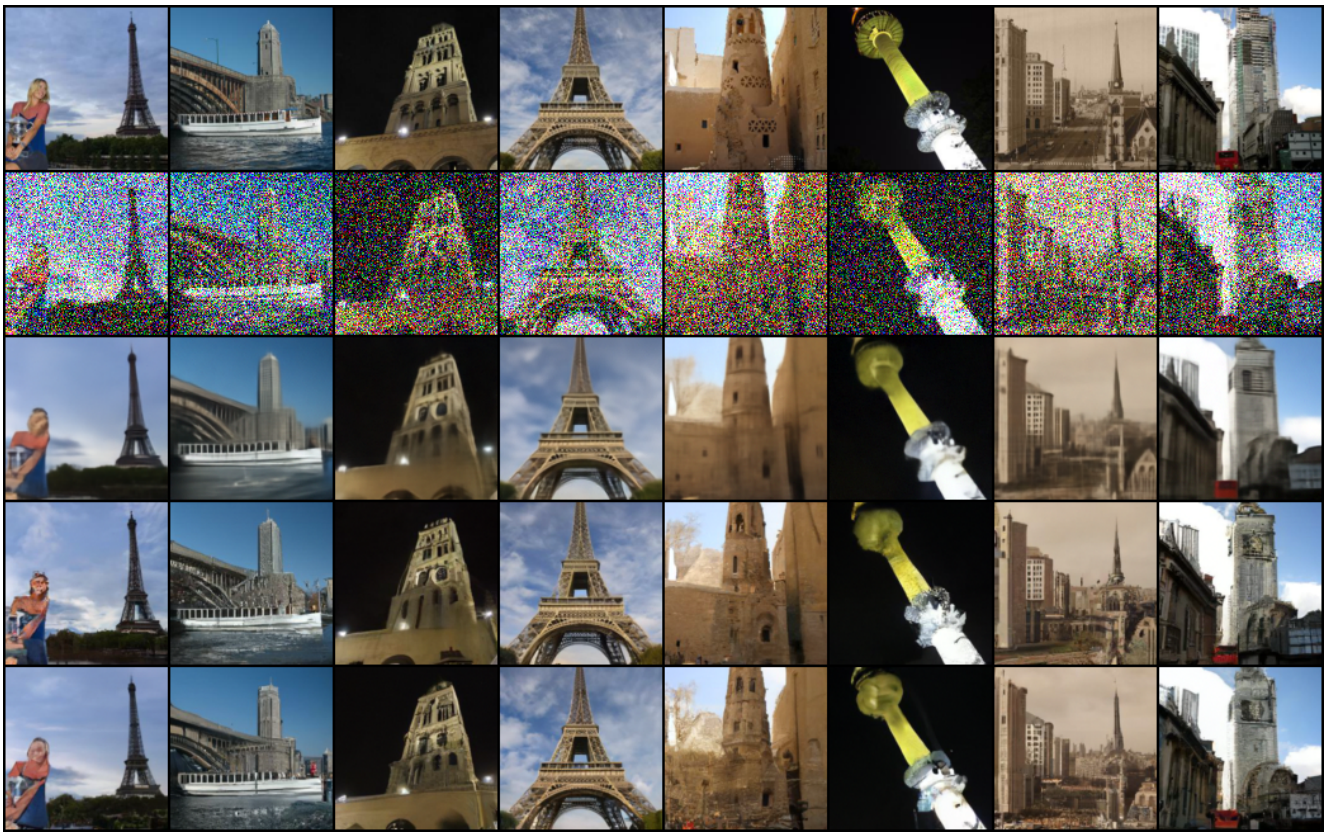

Figure 14. From top to bottom: original LSUN-tower images, noisy versions with $\sigma_0 = 0.403$, MMSE denoiser outputs, and two instances of our denoising algorithm's output.

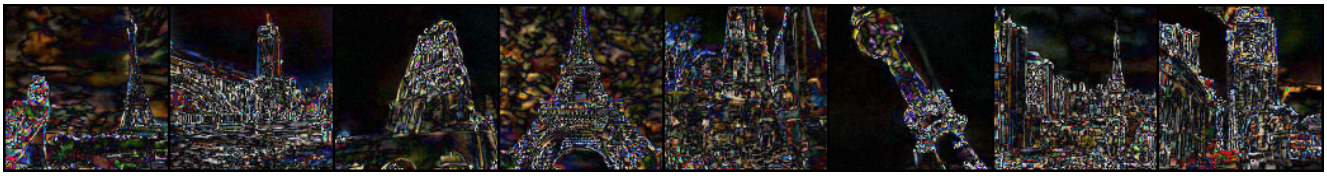

Figure 15. The absolute value of the difference (scaled by 4) between two instances of our denoising algorithm's output on LSUN-tower images with $\sigma_0 = 0.403$.

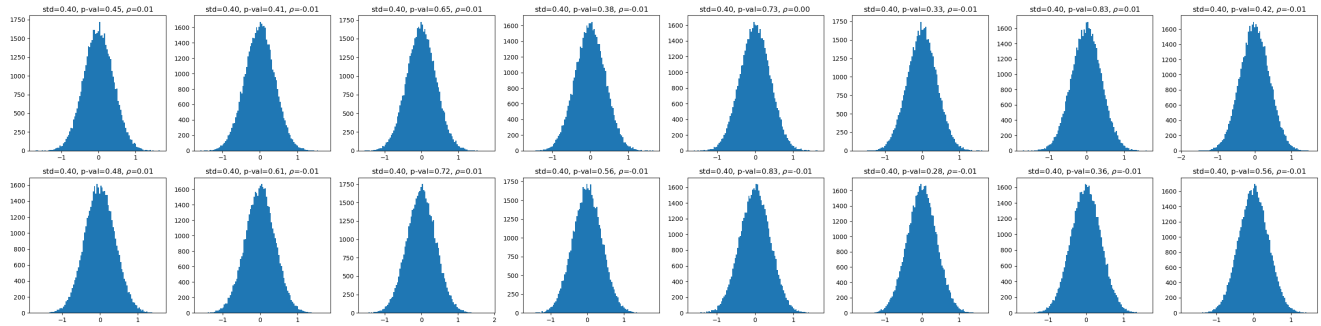

Figure 16. Residual histograms, standard deviations, normality p-values, and the Pearson's correlation coefficients (in the direction with the maximum absolute value) for our denoising algorithm's output on LSUN-tower images with $\sigma_0 = 0.403$.

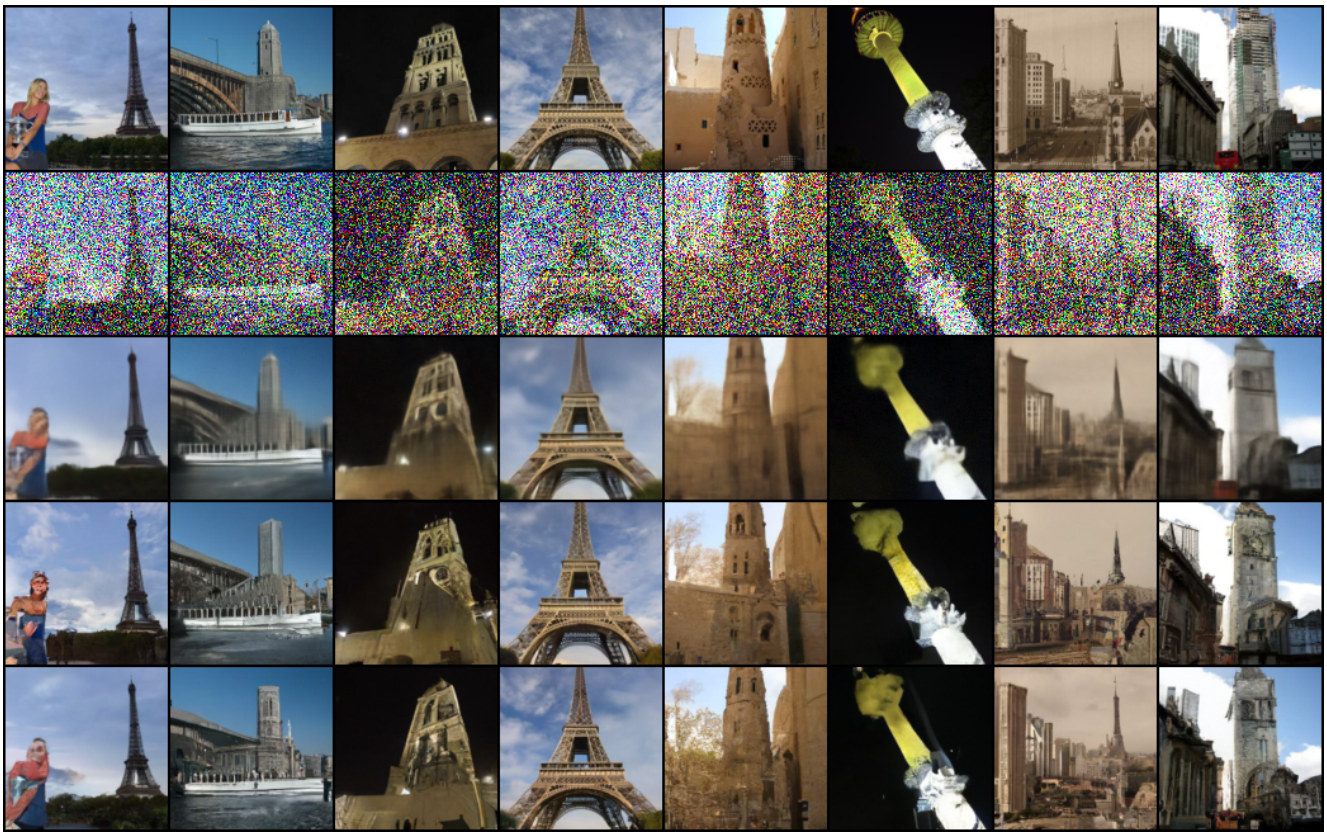

Figure 17. From top to bottom: original LSUN-tower images, noisy versions with $\sigma_0 = 0.606$, MMSE denoiser outputs, and two instances of our denoising algorithm's output.

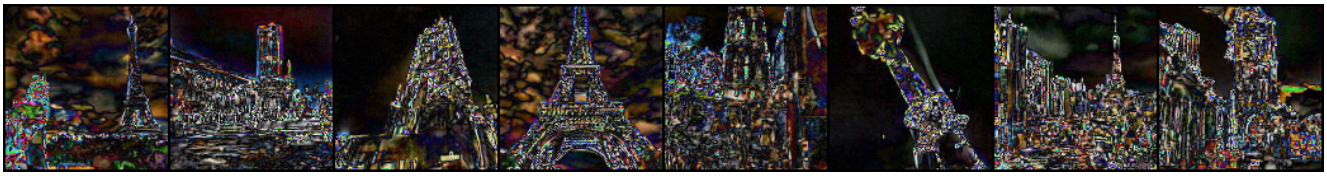

Figure 18. The absolute value of the difference (scaled by 4) between two instances of our denoising algorithm's output on LSUN-tower images with $\sigma_0 = 0.606$.

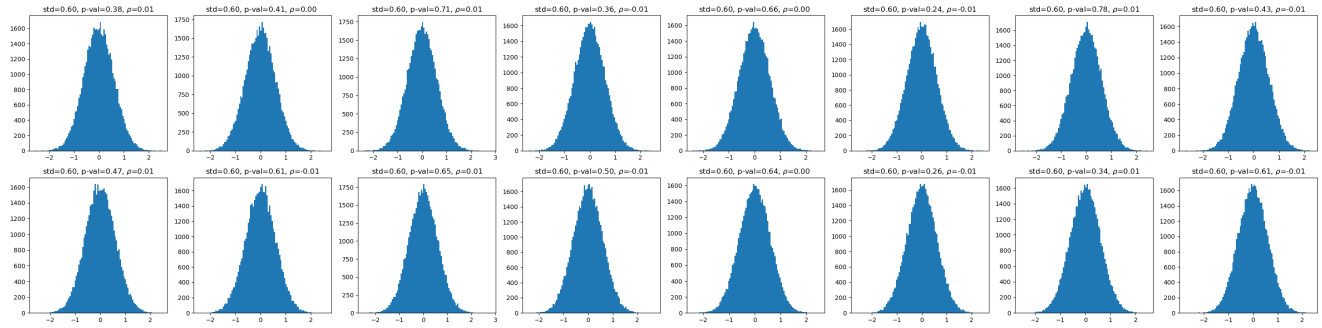

Figure 19. Residual histograms, standard deviations, normality p-values, and the Pearson's correlation coefficients (in the direction with the maximum absolute value) for our denoising algorithm's output on LSUN-tower images with $\sigma_0 = 0.606$.

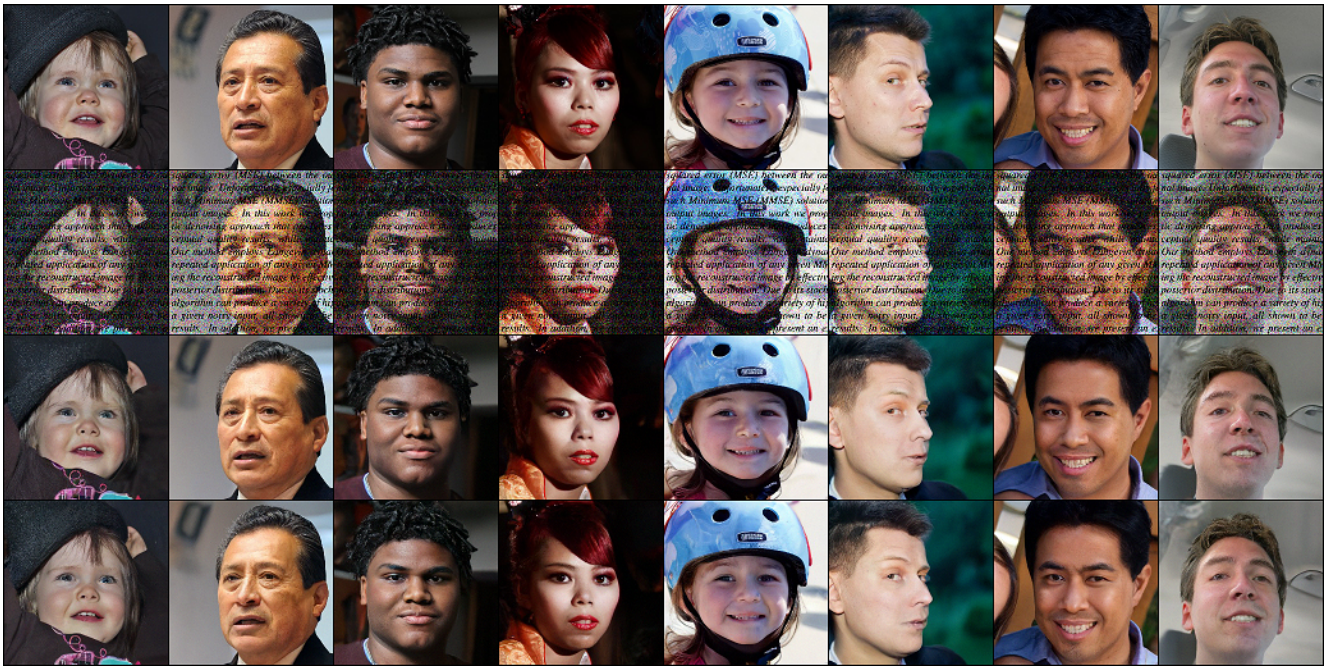

Figure 20. From top to bottom: original FFHQ images, the observations with text overlay and additive noise ($\sigma_0 = 0.2$), and two instances of our inpainting algorithm's output.

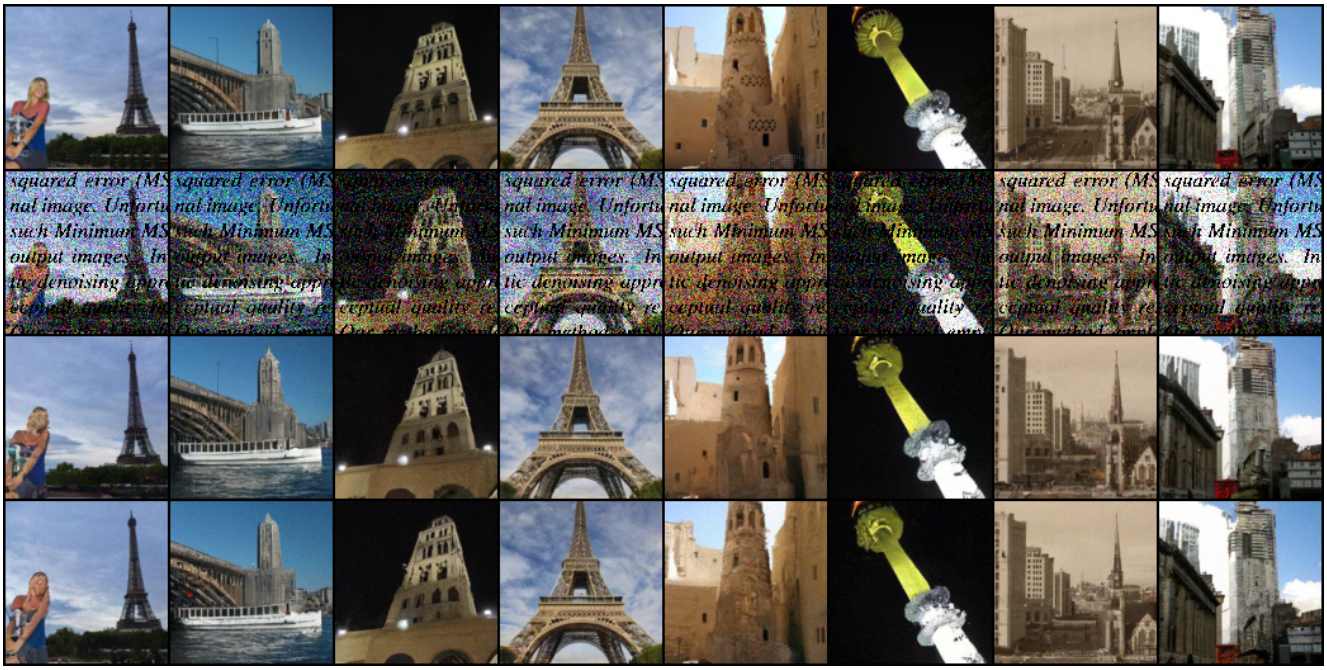

Figure 21. From top to bottom: original LSUN-tower images, the observations with text overlay and additive noise ($\sigma_0 = 0.198$), and two instances of our inpainting algorithm's output.

Figure 22. From top to bottom: original LSUN-bedroom images, the observations with text overlay and additive noise ($\sigma_0 = 0.198$), and two instances of our inpainting algorithm's output.

Figure 23. From top to bottom: original CelebA images, the observations with a missing eye and additive noise ($\sigma_0 = 0.1$), and outputs of our inpainting algorithm.

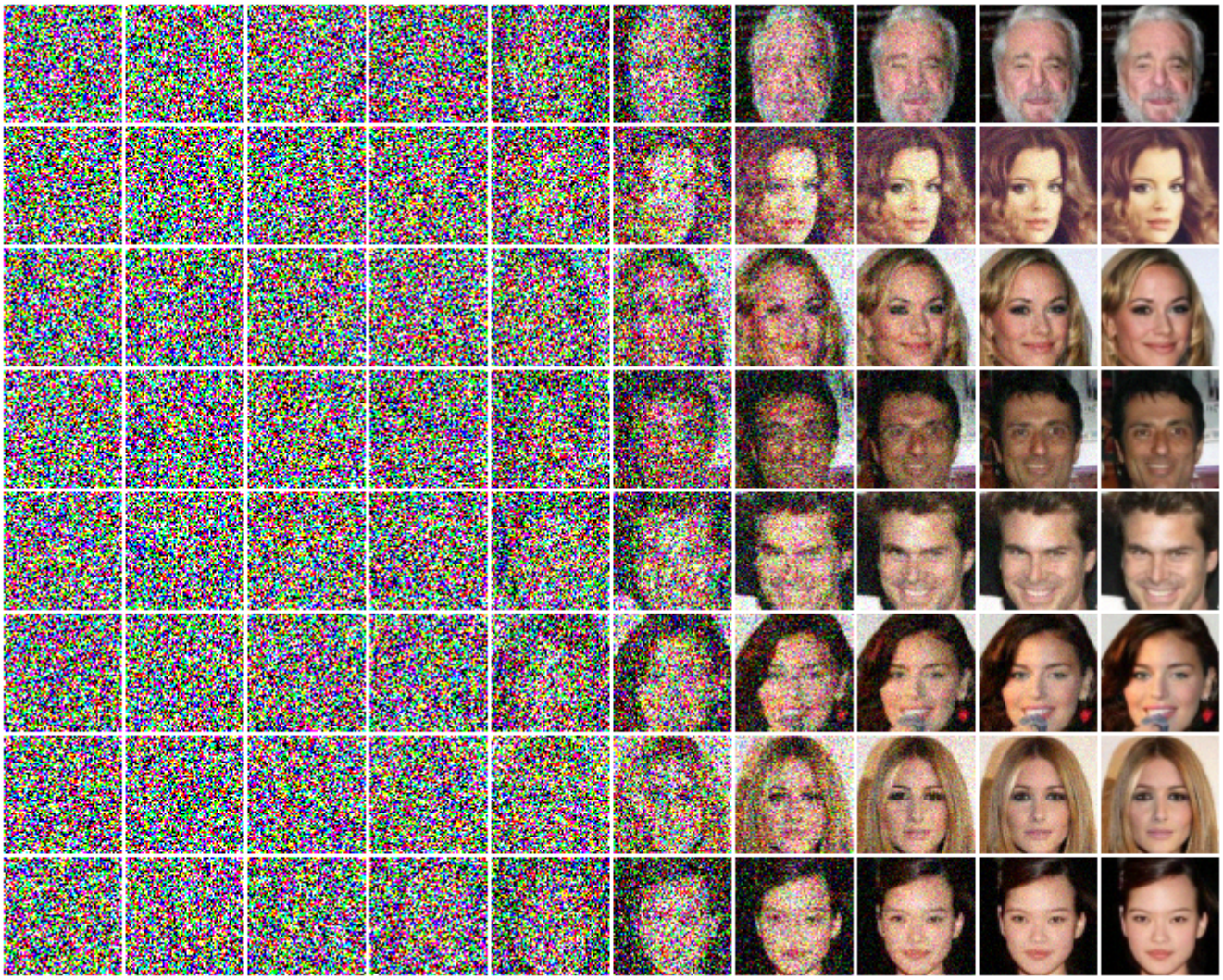

Figure 24. Intermediate results of our inpainting algorithm on CelebA images with a missing eye and additive noise of $\sigma_0 = 0.1$.