Technical Appendix to the paper SDE-based Multiplicative Noise Removal

Anonymous submission

August 16, 2024

1 Solutions to the forward SDE

Consider the 1-dimensional SDE

$$dx = \alpha(t) x d\beta(t) \tag{1.1}$$

This can be solved by applying Itô's formula to $\log x$

$$\begin{aligned} d\log x &= \frac{1}{x}dx - \frac{1}{2x^2}(dx)^2 && (1.2)\\ &= \frac{1}{x}\alpha(t)xd\beta(t) - \frac{1}{2x^2}\alpha^2 x^2 (d\beta(t))^2 && (1.3)\\ &= \alpha(t)d\beta(t) - \frac{1}{2}\alpha^2(t)dt && (1.4)\\ \Rightarrow \quad \log x_t &= \log x_0 - \int_0^t \frac{1}{2}\alpha^2(\tau)d\tau + \int_0^t \alpha(\tau)d\beta(\tau) && (1.5)\end{aligned}$$

Recalling a well-known lemma

Lemma 1. *Let $f(t)$ be some function that is square-integrable, i.e. $\int_0^t f^2(s)ds < \infty$, and $\beta(t)$ be a some Brownian motion, then*

$$\int_0^t f(s)d\beta \sim \mathcal{N}\Big(0, \int_0^t f^2(s)ds\Big) \tag{1.6}$$

Proof. *Since Riemann's sum of $\int_0^t f(s)d\beta$ exists if we fix the midpoints, let $t_k = \frac{k}{2^n}t$, then*

$$\int_0^t f(s)d\beta = \lim_{n\to\infty}\sum_{k=0}^{2^n-1} f(t_k)(\beta_{t_{k+1}} - \beta_{t_k}) \tag{1.7}$$

and since increment of Brownian motion follows $\mathcal{N}(0, \Delta t)$, where $\Delta t = 2^{-n}t$, thus

$$\sum_{k=0}^{2^n-1} f(t_k)(\beta_{t_{k+1}} - \beta_{t_k}) \sim \mathcal{N}\Big(0, \sum_{k=0}^{2^n-1} f^2(t_k)2^{-n}t\Big) \tag{1.8}$$

Now we can take the limit

$$\lim_{n\to\infty}\sum_{k=0}^{2^n-1} f^2(t_k)2^{-n}t = \lim_{n\to\infty}\sum_{k=0}^{2^n-1} f^2(t_k)\Delta t \tag{1.9}$$

$$= \int_0^t f^2(s)ds \tag{1.10}$$

this completes the proof.

Applying this lemma to (1.5) gives

$$\log x_t = \log x_0 - \int_0^t \frac{1}{2}\alpha^2(\tau)d\tau + \Big(\int_0^t \alpha^2(\tau)d\tau\Big)^{\frac{1}{2}} n$$
$$n \sim \mathcal{N}(0,1)$$

Taking exponential of the last equation yields the desired result.

2 Derivations of the reverse SDEs

2.1 In pixel domain

For the reverse SDE, we need to use the more general form of Anderson's theorem

$$dx_{T-t} = \Big(-f(x,T-t) + \frac{1}{p_{T-t}(x_{T-t})}\nabla L^2(x,T-t)p_{T-t}(x_{T-t})\Big)dt + L(x,T-t)d\beta_{T-t}$$

Which in our case simplifies to

$$\begin{aligned} dx_{T-t} &= \alpha^2(T-t)\frac{1}{p_{T-t}(x_{T-t})}\nabla x_{T-t}^2 p_{T-t}(x_{T-t})dt + \alpha(T-t)x_{T-t}d\beta_{T-t} \\ &= \alpha^2(T-t)\Big(2x_{T-t} + x_{T-t}^2\nabla\log p_{T-t}(x_{T-t})\Big)dt + \alpha(T-t)x_{T-t}d\beta_{T-t} \end{aligned} \tag{2.1}$$

This reverse SDE is more complicated to work with because of the spatial dependency of the noise term. Due to this, in order to achieve the same convergence guarantee as Euler-Maruyama discretization, more sophisticated schemes such as Milstein's correction needs to be employed [1], making the generative process much more involved. This motivates us to use the logarithmic formulation, which is shown next.

2.2 In logarithmic domain

Recall that if we apply the logarithmic transform $y = \log x$, then the forward SDE takes the form

$$dy_t = -\frac{1}{2}\alpha^2(t)dt + \alpha(t)d\beta(t) \tag{2.2}$$

which is a simple Wiener process. Applying the previously mentioned Anderson's theorem gives

$$\begin{aligned} dy_{T-t} &= \frac{1}{2}\alpha^2(T-t)dt + \alpha^2(T-t)\frac{1}{p_{T-t}(y_{T-t})}\nabla p_{T-t}(y_{T-t})dt\alpha(T-t)d\beta_{T-t} \\ &= \Big(\frac{1}{2}\alpha^2(T-t) + \alpha^2(T-t)\nabla\log p_{T-t}(y_{T-t})\Big)dt + \alpha(T-t)d\beta_{T-t} \end{aligned} \tag{2.3}$$

Since the corruption applies to each pixel independently, this can be trivially extended to multivariate case to obtain the result mentioned in the paper.

3 Derivation of the deterministic sampling equation using Implicit models

Recall the discretized forward equation presented in the main paper

$$\begin{aligned} \mathbf{y}_k &= \mathbf{y}_{k-1} - \frac{1}{2}\big(\sigma(k) - \sigma(k-1)\big)\mathbf{1} + \sqrt{\sigma(k) - \sigma(k-1)}\mathbf{n}_k \\ &= \mathbf{y}_0 - \frac{1}{2}\big(\sigma(k) - \sigma(0)\big)\mathbf{1} + \sqrt{\sigma(k) - \sigma(0)}\mathbf{n}_k \end{aligned} \tag{3.1}$$

This has the Gaussian transition kernel

$$p(\mathbf{y}_k|\mathbf{y}_{k-1}) = \mathcal{N}\Big(\mathbf{y}_{k-1} - \frac{1}{2}\big(\sigma(k) - \sigma(k-1)\big)\mathbf{1}, \big(\sigma(k) - \sigma(k-1)\big)\mathbf{I}\Big) \tag{3.2}$$

$$p(\mathbf{y}_k|\mathbf{y}_0) = \mathcal{N}\Big(\mathbf{y}_0 - \frac{1}{2}\big(\sigma(k) - \sigma(0)\big)\mathbf{1}, \big(\sigma(k) - \sigma(0)\big)\mathbf{I}\Big) \tag{3.3}$$

Consider the non-Markovian kernel

$$q(\mathbf{y}_{k-1}|\mathbf{y}_k, \mathbf{y}_0) = \mathcal{N}(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \tag{3.4}$$

$$\boldsymbol{\mu}_k = \mathbf{y}_0 - \frac{1}{2}\eta(k-1)\mathbf{1} + \frac{\sqrt{\eta(k-1) - \zeta_k^2}}{\sqrt{\eta(k)}}\big(\mathbf{y}_k - \mathbf{y}_0 + \frac{1}{2}\eta(k)\mathbf{1}\big) \tag{3.5}$$

$$\boldsymbol{\Sigma}_k = \zeta_k^2\mathbf{I} \tag{3.6}$$

where ζ_k^2 is a new parameter controlling the variance of the process, and $\eta(k) = \sigma(k) - \sigma(0)$. We now show that $q(\mathbf{y}_{k-1}|\mathbf{y}_0)$ matches $p(\mathbf{y}_{k-1}|\mathbf{y}_0)$. From [2] (2.115), $q(\mathbf{y}_{k-1}|\mathbf{y}_0)$ is Gaussian $\mathcal{N}(\boldsymbol{\mu}_{k-1}, \boldsymbol{\Sigma}_{k-1})$, and has the following forms

$$\begin{aligned} \boldsymbol{\mu}_{k-1} &= \mathbf{y}_0 - \frac{1}{2}\eta(k-1)\mathbf{1} + \frac{\sqrt{\eta(k-1) - \zeta_k^2}}{\sqrt{\eta(k)}}\big(\mathbf{y}_0 - \frac{1}{2}\eta(k)\mathbf{1} - \mathbf{y}_0 + \frac{1}{2}\eta(k)\mathbf{1}\big) \\ &= \mathbf{y}_0 - \frac{1}{2}\eta(k-1)\mathbf{1} \end{aligned} \tag{3.7}$$

$$\begin{aligned} \boldsymbol{\Sigma}_{k-1} &= \zeta_k^2\mathbf{I} + \Big(\frac{\sqrt{\eta(k-1) - \zeta_k^2}}{\sqrt{\eta(k)}}\Big)^2\eta(k)\mathbf{I} \\ &= \zeta_k^2\mathbf{I} + \big(\eta(k-1) - \zeta_k^2\big)\mathbf{I} \\ &= \eta(k-1)\mathbf{I} \end{aligned} \tag{3.8}$$

Thus, $q(\mathbf{y}_{k-1}|\mathbf{y}_0)$ has the Gaussian kernel

$$\begin{aligned} q(\mathbf{y}_{k-1}|\mathbf{y}_0) &= \mathcal{N}(\boldsymbol{\mu}_{k-1}, \boldsymbol{\Sigma}_{k-1}) \\ &= \mathcal{N}\Big(\mathbf{y}_0 - \frac{1}{2}\eta(k-1)\mathbf{1}, \eta(k-1)\mathbf{I}\Big) \\ &= \mathcal{N}\Big(\mathbf{y}_0 - \frac{1}{2}\big(\sigma(k-1) - \sigma(0)\big)\mathbf{1}, \big(\sigma(k-1) - \sigma(0)\big)\mathbf{I}\Big) \\ &= p(\mathbf{y}_{k-1}|\mathbf{y}_0) \end{aligned} \tag{3.9}$$

This completes the proof.

4 Additional experiment results

We provide more test samples in Figures 1 and 2. It can be seen that our method tends to preserver high-frequency components better, while other approaches produce over-smoothed images.

5 Reproducibility Checklist

This paper:

- Includes a conceptual outline and/or pseudocode description of AI methods introduced (**yes**/partial/no/NA)
- Clearly delineates statements that are opinions, hypothesis, and speculation from objective facts and results (**yes**/no)
- Provides well marked pedagogical references for less-familiar readers to gain background necessary to replicate the paper (**yes**/no)

Does this paper make theoretical contributions? (yes/**no**)
Does this paper rely on one or more datasets? (**yes**/no)

- A motivation is given for why the experiments are conducted on the selected datasets (**yes**/partial/no/NA)
- All novel datasets introduced in this paper are included in a data appendix. (yes/partial/no/**NA**)
- All novel datasets introduced in this paper will be made publicly available upon publication of the paper with a license that allows free usage for research purposes. (yes/partial/no/**NA**)
- All datasets drawn from the existing literature (potentially including authors' own previously published work) are accompanied by appropriate citations. (**yes**/no/NA)
- All datasets drawn from the existing literature (potentially including authors' own previously published work) are publicly available. (**yes**/partial/no/NA)
- All datasets that are not publicly available are described in detail, with explanations why publicly available alternatives are not scientifically satisfactory. (yes/partial/no/**NA**)

Does this paper include computational experiments? (**yes**/no)

- Any code required for pre-processing data is included in the appendix. (**yes**/partial/no).
- All source code required for conducting and analyzing the experiments is included in a code appendix. (**yes**/partial/no)

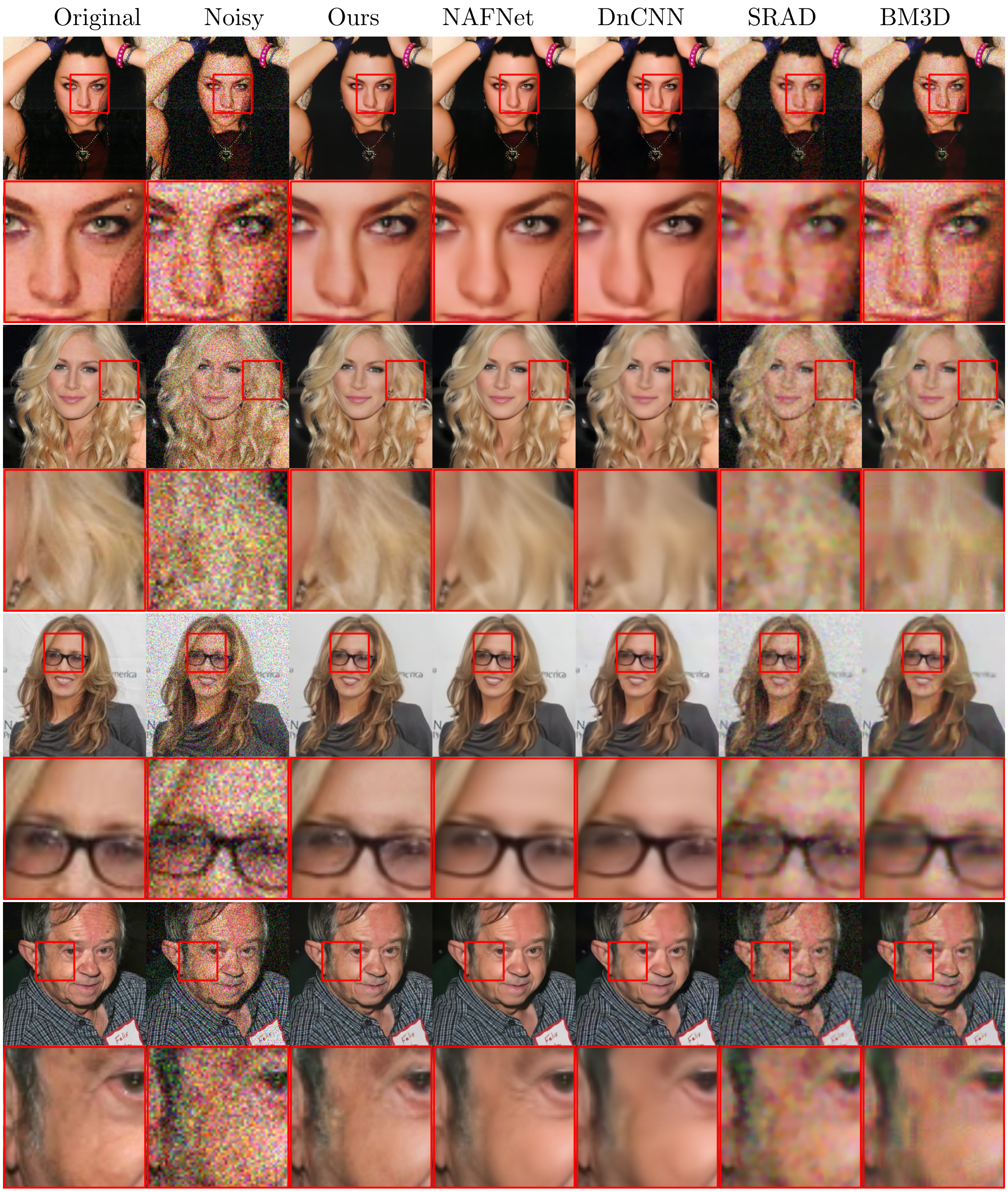

Figure 1: Comparing between different denoising models on randomly selected CelebA images, at noise level 0.12. The first two columns include the original images and their noised versions, respectively. These are followed by the results generated by our method and other popular techniques. Full resolution version is available in supplemental materials.

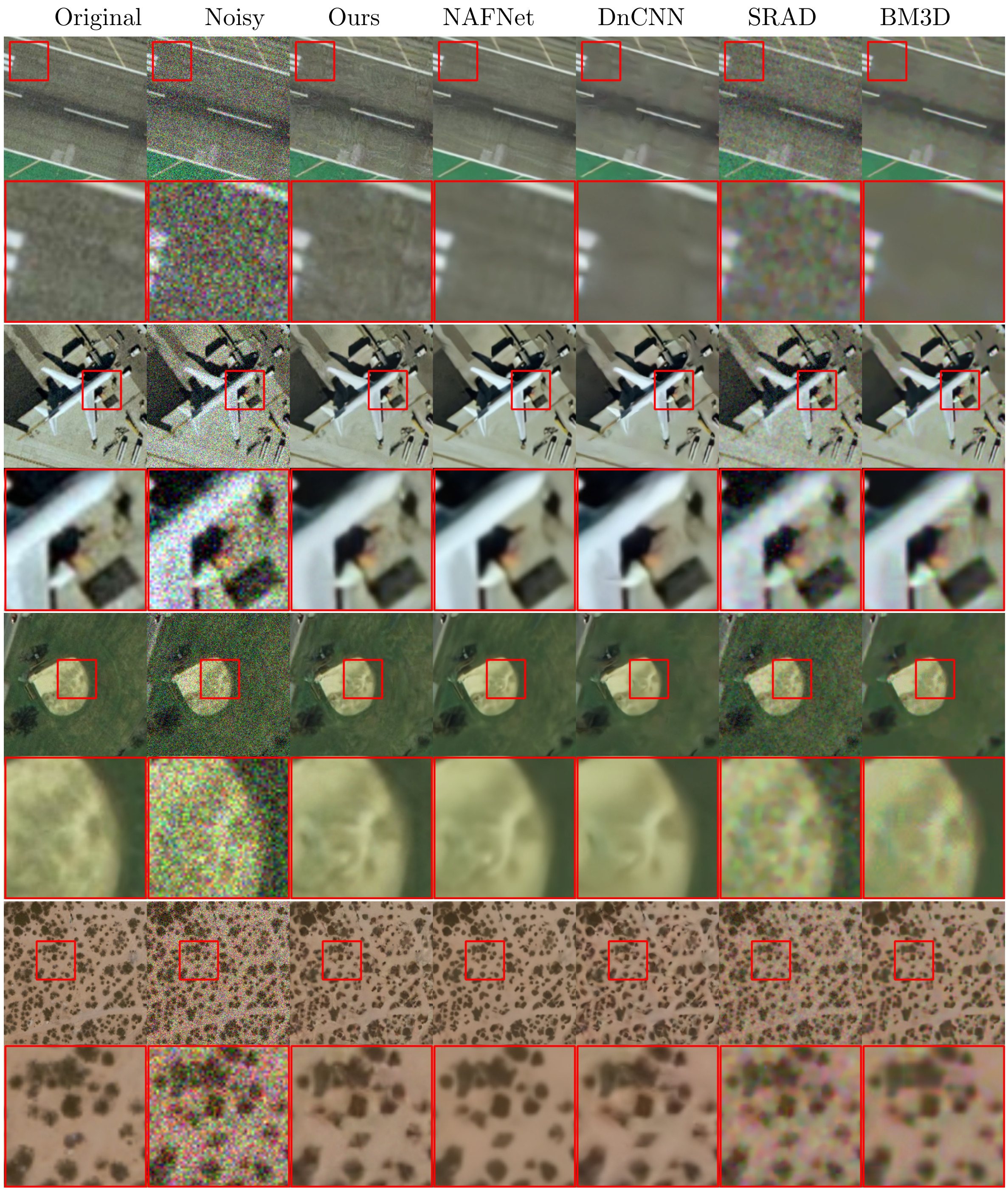

Figure 2: Comparing between different denoising models on randomly selected UC Merced Land Use images, at noise level 0.12. The first two columns include the original images and their noised versions, respectively. These are followed by the results generated by our method and other popular techniques. Full resolution version is available in supplemental materials.

- All source code required for conducting and analyzing the experiments will be made publicly available upon publication of the paper with a license that allows free usage for research purposes. (**yes**/partial/no)
- All source code implementing new methods have comments detailing the implementation, with references to the paper where each step comes from (yes/partial/no)
- If an algorithm depends on randomness, then the method used for setting seeds is described in a way sufficient to allow replication of results. (**yes**/partial/no/NA)
- This paper specifies the computing infrastructure used for running experiments (hardware and software), including GPU/CPU models; amount of memory; operating system; names and versions of relevant software libraries and frameworks. (yes/**partial**/no)
- This paper formally describes evaluation metrics used and explains the motivation for choosing these metrics. (**yes**/partial/no)
- This paper states the number of algorithm runs used to compute each reported result. (**yes**/no)
- Analysis of experiments goes beyond single-dimensional summaries of performance (e.g., average; median) to include measures of variation, confidence, or other distributional information. (yes/**no**)
- The significance of any improvement or decrease in performance is judged using appropriate statistical tests (e.g., Wilcoxon signed-rank). (yes/partial/**no**)
- This paper lists all final (hyper-)parameters used for each model/algorithm in the paper's experiments. (**yes**/partial/no/NA)
- This paper states the number and range of values tried per (hyper-) parameter during development of the paper, along with the criterion used for selecting the final parameter setting. (yes/partial/no/**NA**)

References

[1] P. E. Kloeden, E. Platen, P. E. Kloeden, and E. Platen, *Stochastic differential equations*. Springer, 1992.

[2] C. M. Bishop and N. M. Nasrabadi, *Pattern recognition and machine learning*. Springer, 2006, vol. 4.