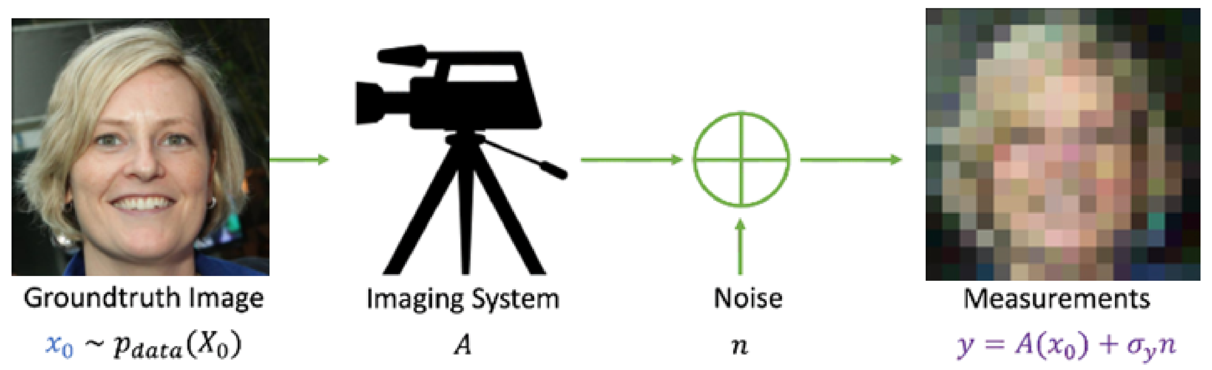
Towards Frugal Zero-Shot Diffusion Based Image Restoration

Rajit Rajpal [1] Marcos Obando [2] Dolly Chen [3] Bernardin Tamo Amougou [4]

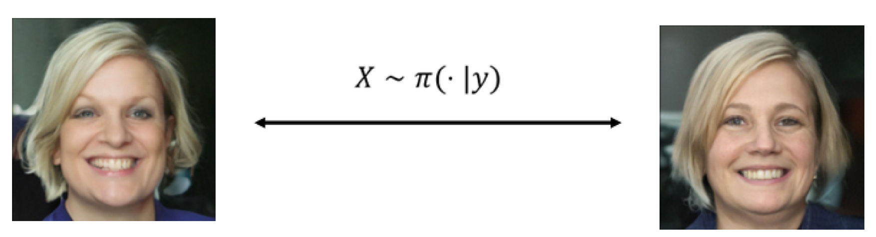
[1]University of Edinburgh [2]CWI [3]UCL [4]Heriot-Watt University

Introduction to Inverse Problems



Problem: Reconstruct groundtruth image $\mathbf{x}_0$ from noisy measurements y

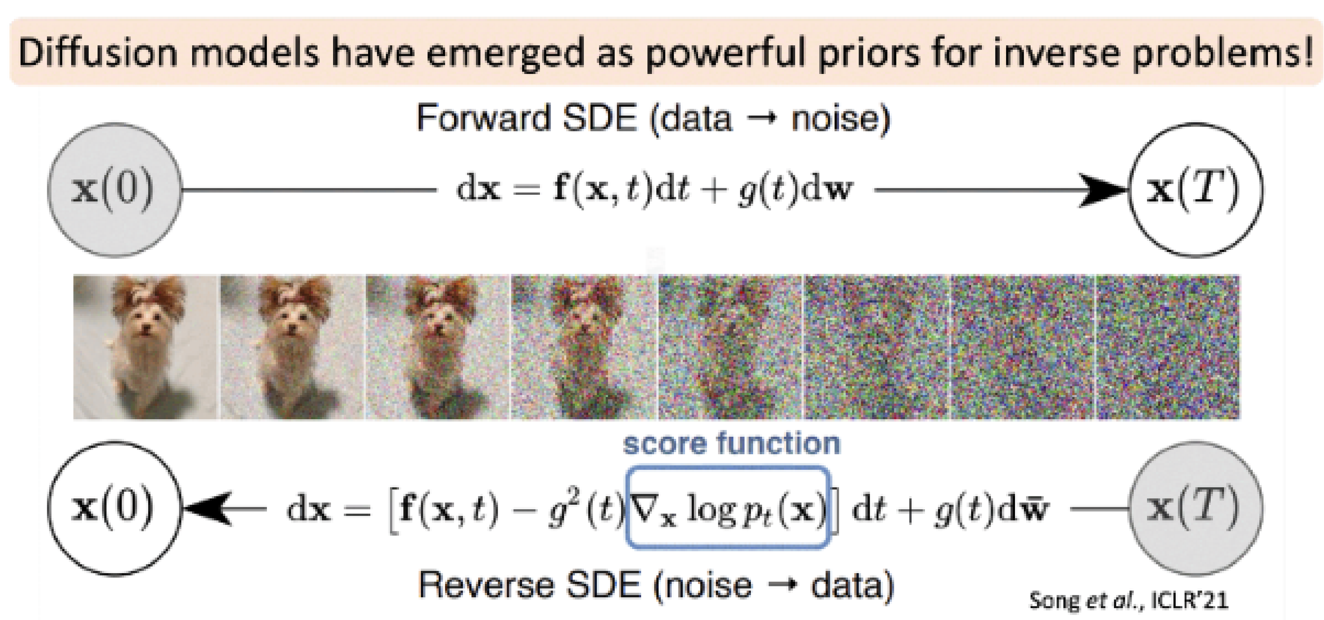
Challenge: Problem is ill-posed, that is infinitely many solutions $\mathbf{x}_0$ exist



Approach: Use prior knowledge $p(\mathbf{x}_0)$ of how the image should look like

Applications: Denoising, Colorization, Superresolution, Deblurring, Inpainting, Outpainting

Background on Diffusion Models



Posterior Sampling Using Diffusion Models

Problem: Sample $p_0(\mathbf{x}_0 \mid \mathbf{y})$ instead of $p(\mathbf{x}_0)$

$$d\mathbf{x}_t = (-\mathbf{x}_t - 2\nabla \log p_t(\mathbf{x}_t \mid y))\, dt + \sqrt{2}d\bar{W}_t, \quad t = T, \cdots, 0$$

$$d\mathbf{x}_t = (-\mathbf{x}_t - 2\nabla \log p_t(\mathbf{y} \mid \mathbf{x}_t) - 2\nabla \log p_t(\mathbf{x}_t))\, dt + \sqrt{2}d\bar{W}_t \quad \text{(Bayes rule)}$$

Unknown: $\nabla \log p_t(y \mid \mathbf{x}_t)$ Known: $\nabla \log p_t(\mathbf{x}_t) \approx s_\theta(\mathbf{x}_t, t)$

Approach: Sample $p_0(\mathbf{x}_0 \mid \mathbf{y})$ using $\nabla \log p(\mathbf{x}_t)$ without having to re-train:

1. a new score function $\nabla \log p_t(\mathbf{x}_t \mid \mathbf{y})$ or
2. a noise-conditional measurement model : $\nabla \log p_t(\mathbf{y} \mid \mathbf{x}_t)$

How well can we approximate $\nabla \log p_t(y|x_t)$?

Methodology

To have a DDM for the posterior $\pi(.|\mathbf{y})$, we need a parametric approximation of

$$\mathbb{E}[\mathbf{x}_0|\mathbf{x}_k = x_k, \mathbf{y} = \mathbf{y}] = \int \mathbf{x}_0\, q_{0|k}(\mathbf{x}_0|\mathbf{x}_k, \mathbf{y})\, d\mathbf{x}_0, \tag{1}$$

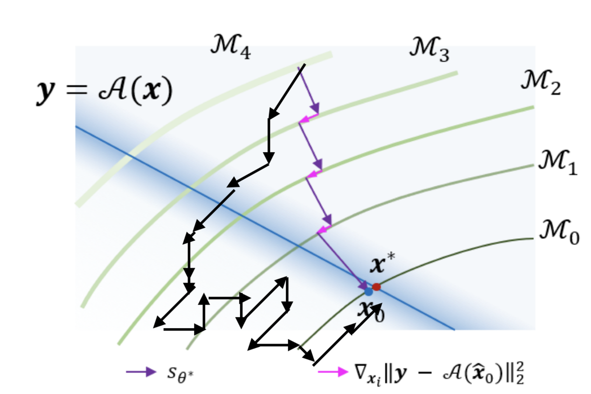
where $q_{0|k}(\mathbf{x}_0|\mathbf{x}_k, \mathbf{y}) \propto \pi(\mathbf{x}_0|\mathbf{y}) q_{k|0}(\mathbf{x}_k|\mathbf{x}_0)$ and $0 \leq k \leq t$.

Via Tweedie's formula, one can show that [4]

$$\mathbb{E}[\mathbf{x}_0|\mathbf{x}_k, \mathbf{y}] = \mathbb{E}[\mathbf{x}_0|\mathbf{x}_k] + \frac{1 - \bar{\alpha}_k}{\bar{\alpha}_k^{1/2}} \nabla_{\mathbf{x}_k} \log \int p(\mathbf{y}|\mathbf{x}_0)\, q_{0|k}(\mathbf{x}_0|\mathbf{x}_k)\, d\mathbf{x}_0 \tag{2}$$

1. [2] proposed **Diffusion Posterior Sampling (DPS)** method:
$$\nabla_k \log \int p(\mathbf{y}|\mathbf{x}_0)\, q_{0|k}(\mathbf{x}_0|\mathbf{x}_k)\, d\mathbf{x}_0 \approx \nabla_{\mathbf{x}_k} \log p(\mathbf{y}|\mathbb{E}[\mathbf{x}_0|\mathbf{x}_k]), \tag{3}$$
which implicitly assumes that $q_{0|k}(\cdot|\mathbf{x}_k) \approx \delta_{\mathbf{x}^\theta_{0|k}(\mathbf{x}_k)}$.
2. [3] propose **Diffusion Model Based Posterior Sampling (DMPS)** which claims to be 3x faster per function evaluation that DPS. They make the uninformative prior assumption:
$$p(\mathbf{x}_0|\mathbf{x}_t) \propto p(\mathbf{x}_t|\mathbf{x}_0) = \mathcal{N}(\mathbf{x}_0; \frac{\mathbf{x}_t}{\sqrt{\bar{\alpha}_t}}, \frac{1 - \bar{\alpha}_t}{\bar{\alpha}_t}\mathbf{I}). \tag{4}$$
This allows us to form an analytical expression for the second term of Eq 2.
$$\nabla_{\mathbf{x}_k} \log \int p(\mathbf{y}|\mathbf{x}_0)\, q_{0|k}(\mathbf{x}_0|\mathbf{x}_k)\, d\mathbf{x}_0 \approx \frac{1}{\sqrt{\bar{\alpha}}}\mathbf{A}^T(\sigma^2\mathbf{I} + \frac{1 - \bar{\alpha}_k}{\bar{\alpha}_k}\mathbf{A}\mathbf{A}^T)^{-1}(\mathbf{y} - \frac{1}{\sqrt{\bar{\alpha}_k}}\mathbf{A}x_k) \tag{5}$$
$$= \frac{1}{\sqrt{\bar{\alpha}}}\mathbf{V}\mathbf{\Sigma}(\sigma^2\mathbf{I} + \frac{1 - \bar{\alpha}_k}{\bar{\alpha}_k}\mathbf{\Sigma}^2)^{-1}(\mathbf{U}^T y - \frac{1}{\sqrt{\bar{\alpha}_k}}\mathbf{\Sigma}\mathbf{V}^T x_k) \tag{6}$$
The SVD in Eq 6 circumvents inversion of $\mathbf{A}$. It only needs to be computed once.

DMPS makes a crude assumption, in that it implicitly assumes that $\frac{p(\mathbf{x_0})}{p(\mathbf{x_k})} \propto 1$. For this reason, [3] were unable to perform difficult tasks such as inpainting and outpainting.



$$\hat{\mathbf{x}}_0(\mathbf{x}_k) = \frac{\mathbf{x}_k}{\sqrt{\bar{\alpha}_k}} - \sqrt{\frac{1 - \bar{\alpha}}{\bar{\alpha}}} \begin{cases} \mathbf{s}_\theta(\mathbf{x_k}, k) & \text{DPS} \\ \mathbf{w} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}) & \text{DMPS} \end{cases} \tag{7}$$

The equation above indicates that the samples $\mathbf{x}_0|\mathbf{x}_k$ for DPS are driven by the score $\mathbf{s}_\theta(\mathbf{x_k}, k)$ highlighting exploitation, whereas DMPS is driven by $\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ relies on exploration. The black arrows in the figure above [2] is an instance where it chances upon the optimal generative manifold $\mathbf{y} = \mathcal{A}(x^*)$. This is not guaranteed in the case where the prior $p_\theta(x)$ and posterior $\pi(.|\mathbf{y})$ are significantly different e.g. inpainting.

Idea

Use DPS to sample from time t to k and DMPS to sample from time k to 0.

Motivation: [1] theorizes that the generative dynamics of diffusion models starts from pure noise and encounters first a 'speciation' transition at time t_S where the gross structure of data is unraveled, and then a 'collapse' transition at time t_C where the trajectories of the dynamics become attracted to one of the memorized data points. $\therefore$at t_C, it is reasonable to assume $p(\mathbf{x}_0|\mathbf{x}_k) \propto p(\mathbf{x}_k|\mathbf{x}_0)$ for $k << t$ i.e., $k = t_C$. We develop a heuristic to determine t_C by finding k for which: pSNR$(\mathbf{x}_0|\mathbf{x}_k - \mathbf{x}_0|\mathbf{x}_{k-1}) \geq 20$, which we determined empirically for CelebA.

Numerical Experiments

We used a pre-trained CelebA diffusion model and evaluated on tasks that [3] did not.

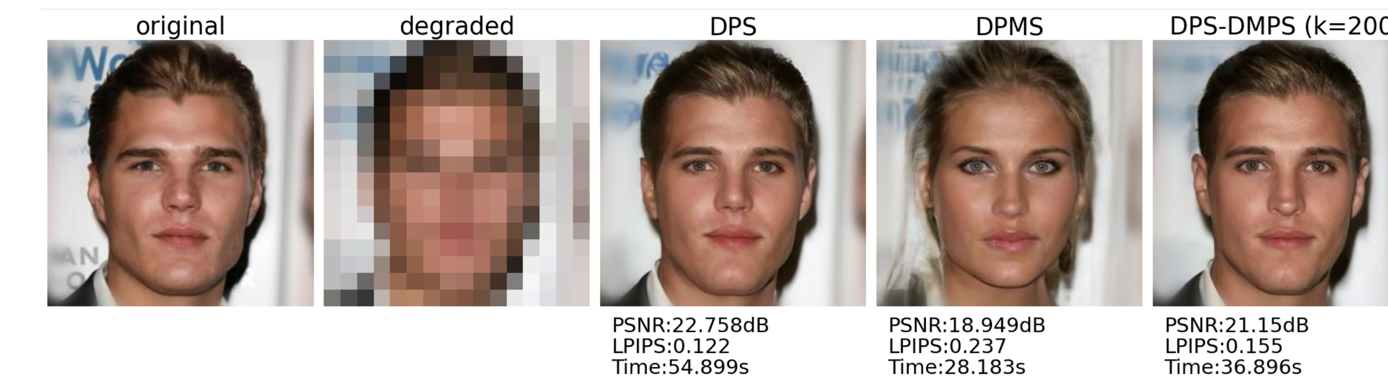


Figure 1. Super-resolution x16, using $\sigma = 0.01$, $t = 300$.

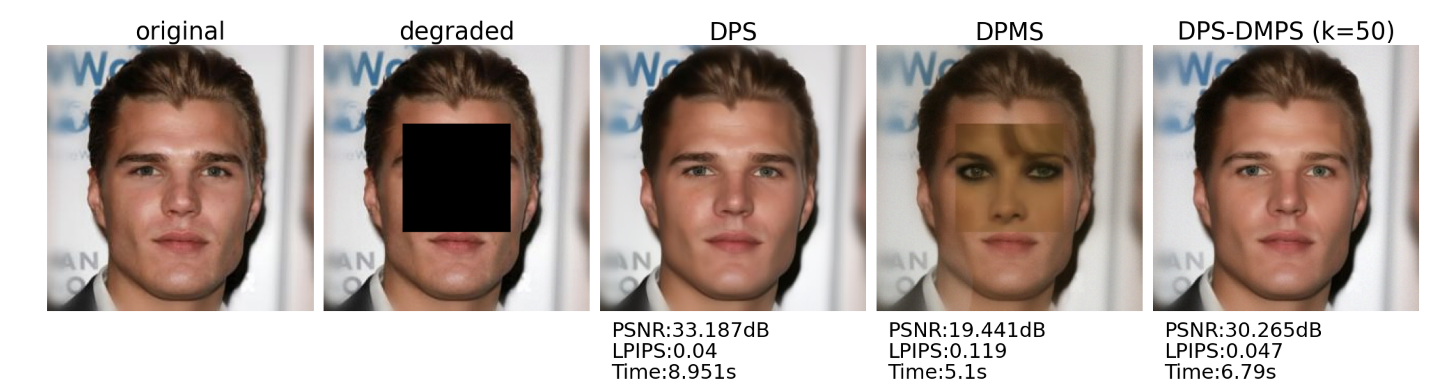


Figure 2. Inpainting, using $\sigma = 0.01$, $t = 100$.

	Time (s)↓	pSNR (db)↑	LPIPS↓
DPS	17.72	19.37	0.25
DMPS	**9.19**	15.63	0.41
DPS+DMPS	13.42	**19.97**	**0.24**

Table 1. SR-16 metrics averaged over 20 in-distribution images for $t = 100$, $k = 50$, $\sigma = 0.01$

We achieve, on average, better metrics on in-distribution images than DPS in a shorter time.

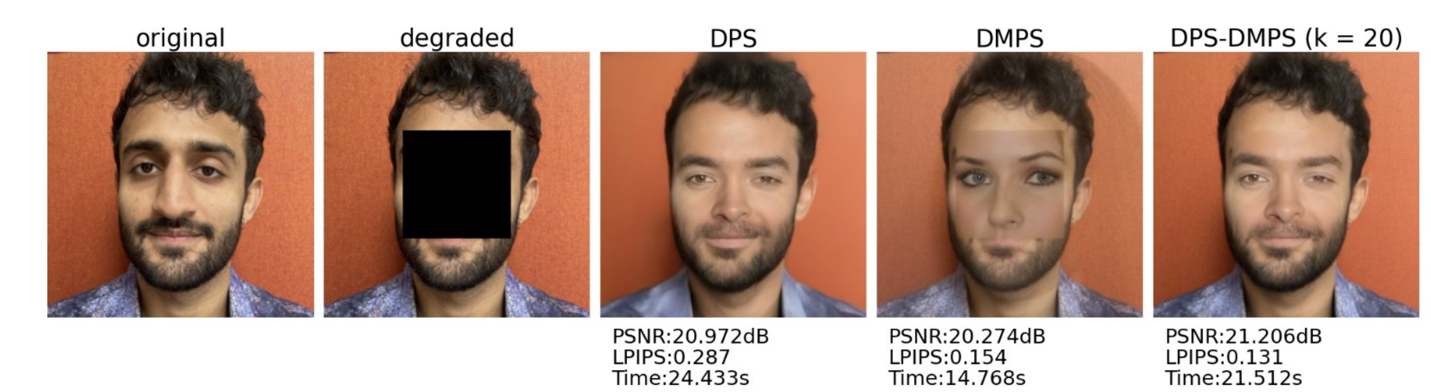


Figure 3. OOD inpainting using $\sigma = 0.01$, $t = 100$.

As indicated by the out-of-distribution task in Fig 3, our method is more robust (lower LPIPS) than DPS because it promotes exploration once a mode is found (escaping sharp local minima).

References

[1] Giulio Biroli, Tony Bonnaire, Valentin de Bortoli, and Marc Mézard. Dynamical regimes of diffusion models, 2024.

[2] Hyungjin Chung, Jeongsol Kim, Michael T. Mccann, Marc L. Klasky, and Jong Chul Ye. Diffusion posterior sampling for general noisy inverse problems, 2024.

[3] Xiangming Meng and Yoshiyuki Kabashima. Diffusion model based posterior sampling for noisy linear inverse problems, 2024.

[4] Yang Song, Jascha Sohl-Dickstein, Diederik P. Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations, 2021.

[1]R.Rajpal@sms.ed.ac.uk, [2]Marcos.Obando@cwi.nl, [3]dai.chen.22@ucl.ac.uk, [4]bt2027@hw.ac.uk