Inverting scientific images with score-based generative models

September 27, 2024

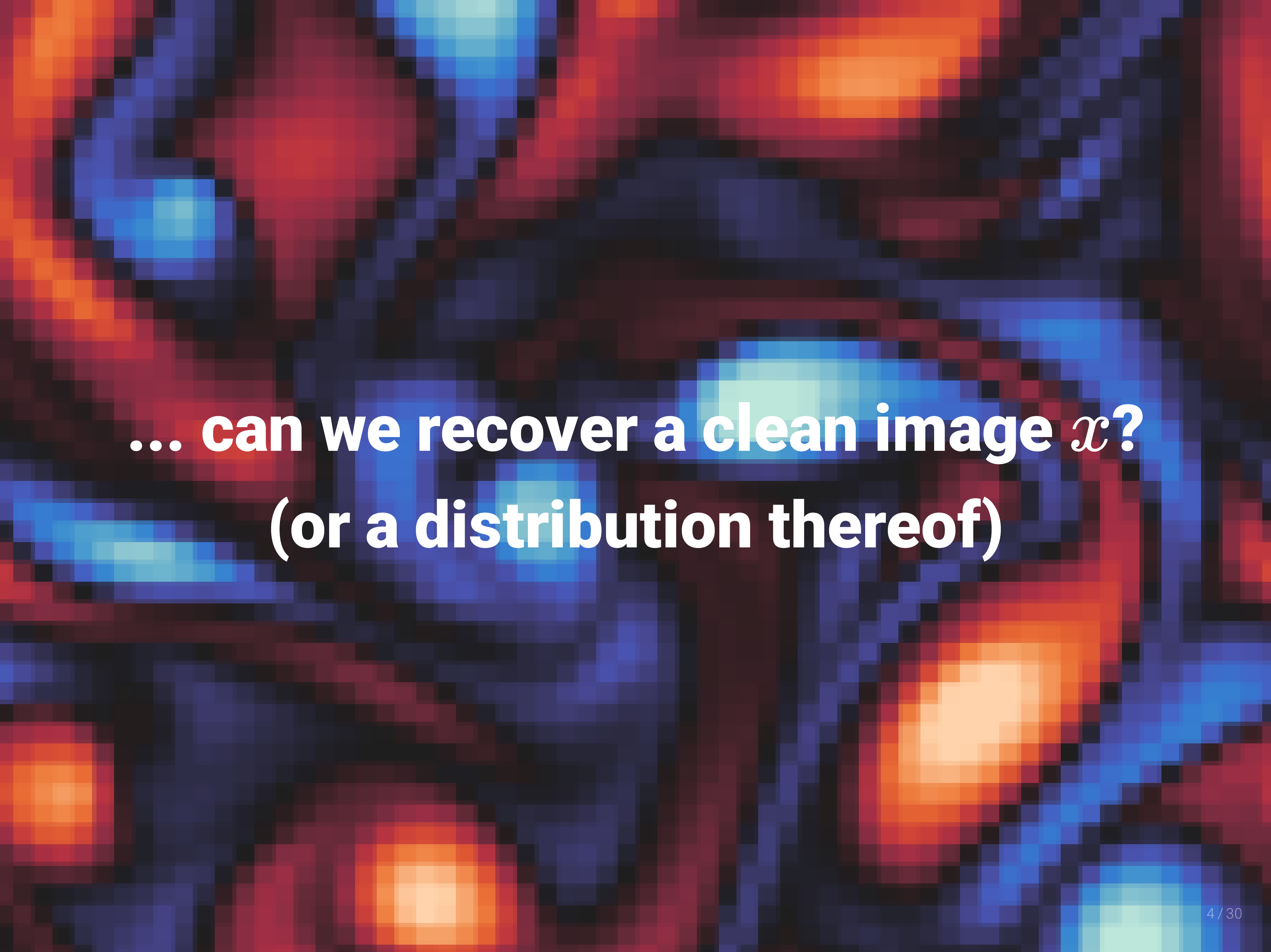
Gilles Louppe
g.louppe@uliege.be

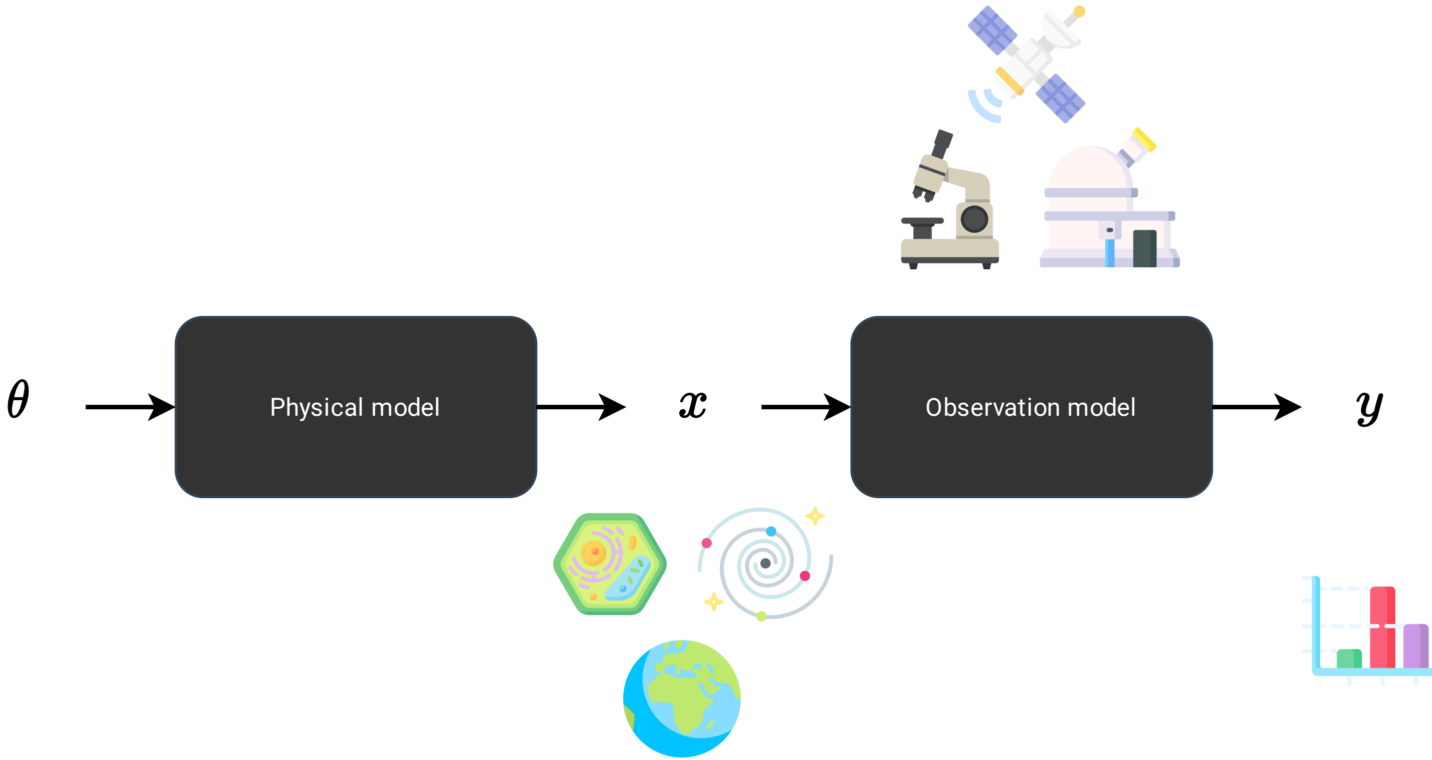
LIÈGE université

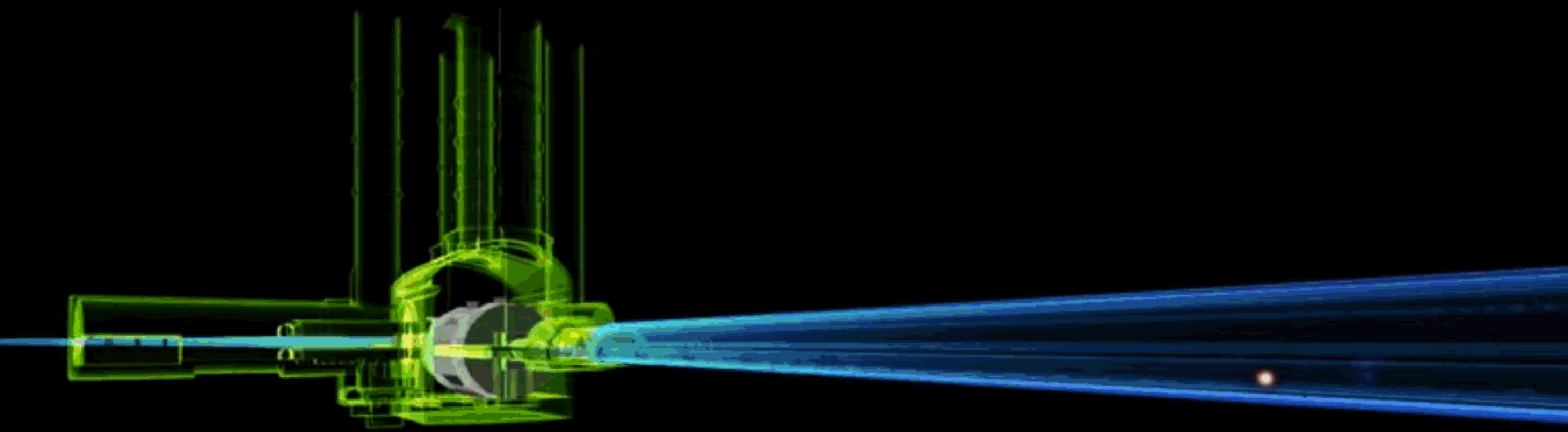
References:

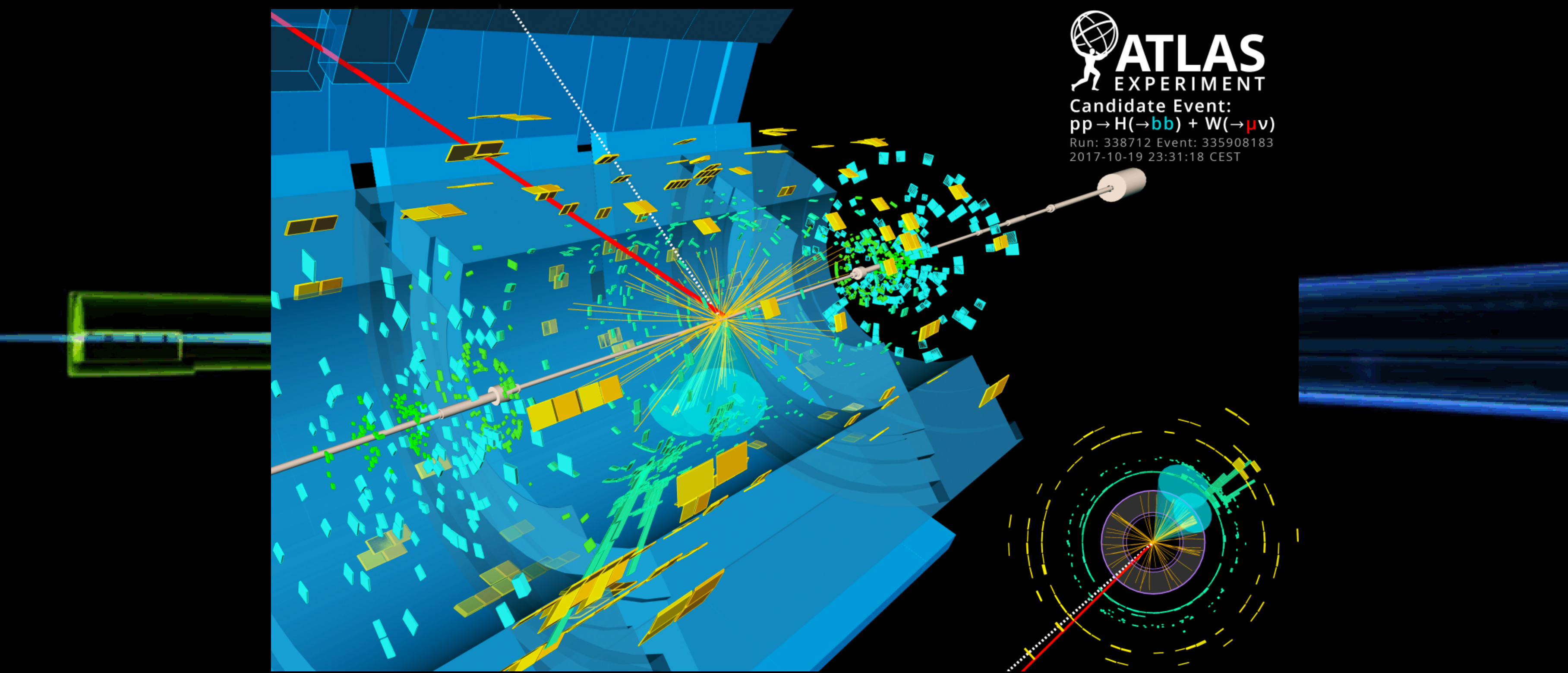
- **Score-based data assimilation**
  François Rozet, Gilles Louppe.
  NeurIPS 2023, arXiv:2306.10574.
- **Score-based Data Assimilation for a Two-Layer Quasi-Geostrophic Model**
  François Rozet, Gilles Louppe.
  ML4PS workshop NeurIPS 2023, arXiv:2310.01853.
- **Learning Diffusion Priors from Observations by Expectation Maximization**
  François Rozet, Gérôme Andry, François Lanusse, Gilles Louppe.
  NeurIPS 2024, arXiv:2405.13712.

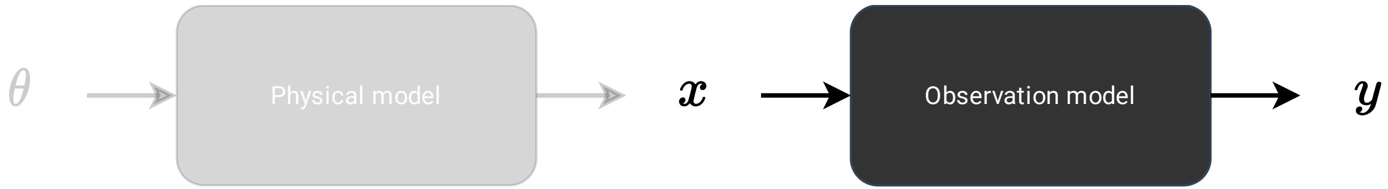
From a noisy observation y...


... can we recover a clean image x?
(or a distribution thereof)


θ
Physical model
x
Observation model
y




ATLAS
EXPERIMENT
Candidate Event:
pp→H(→bb) + W(→μν)
Run: 338712 Event: 335908183
2017-10-19 23:31:18 CEST



Problem statement

Estimate the latent state x from the observation y through the Bayesian posterior

$$p(x|y) = \frac{p(y|x)p(x)}{p(y)},$$

where

- $p(y|x)$ is a known observation model,
- $p(x)$ is a prior distribution over the latent state,
- $p(y)$ is the marginal density of the observations.

Note: In this talk, we do not consider the inference of model parameters θ given y.

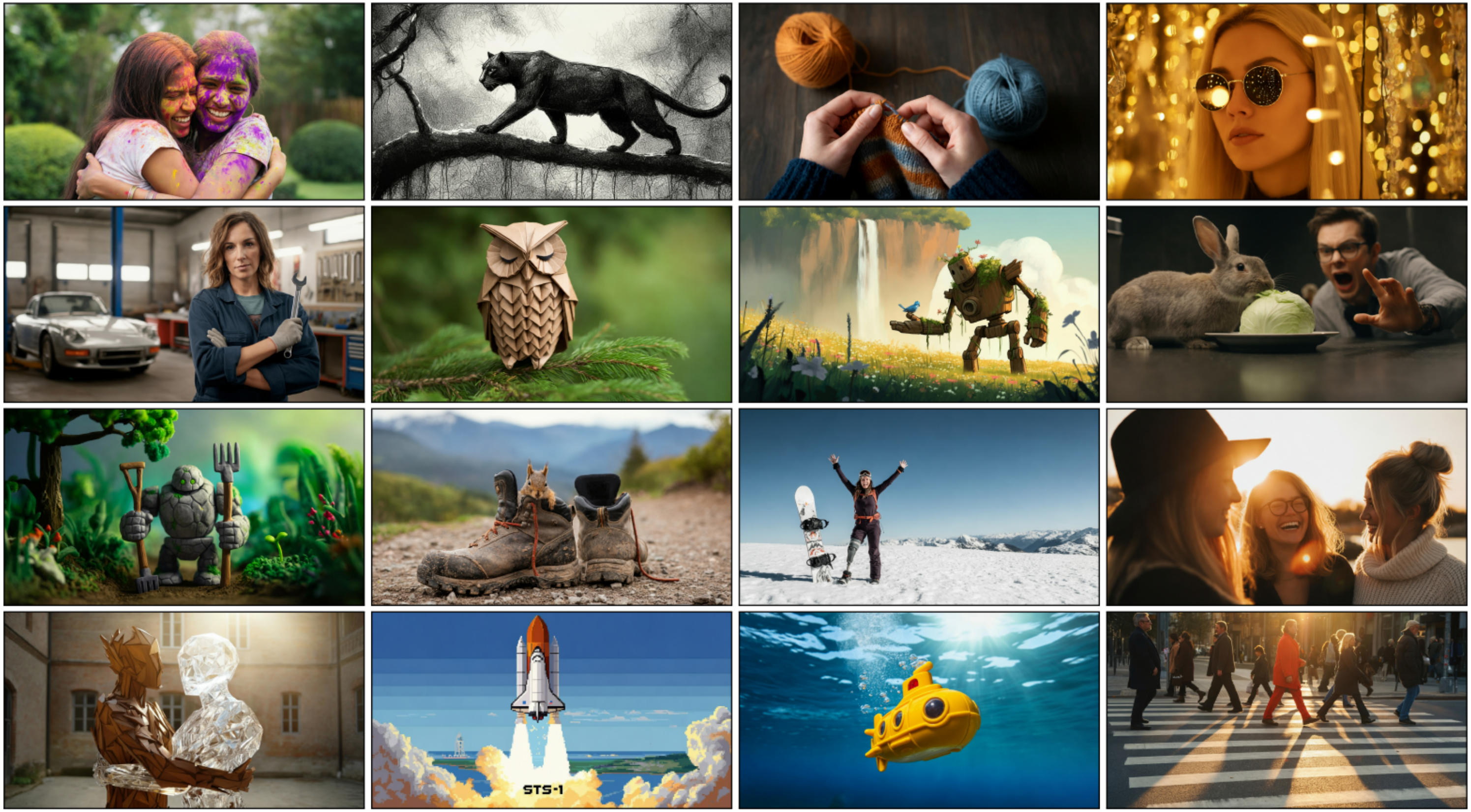
Simulation-based inference

Neural network surrogates $q(x|y)$ of the posterior $p(x|y)$ can be trained in various ways, for example by using a conditional density estimator $q(x|y)$ and directly maximizing

$$\mathbb{E}_{p(y)p(x|y)}\left[\log q(x|y)\right].$$

Issues:

- Neural density estimators, such as conditional normalizing flows, do not scale well to high dimensions (e.g., when x is an image).
- The neural surrogate $q(x|y)$ is wired to the observation model $p(y|x)$, and must be retrained when it changes.
- Paired data (x, y) are required for training.



Proposition: Score-based generative models can address all those issues.

—
Credits: Imagen 3 Team, Google, 2024 (arXiv:2408.07009).

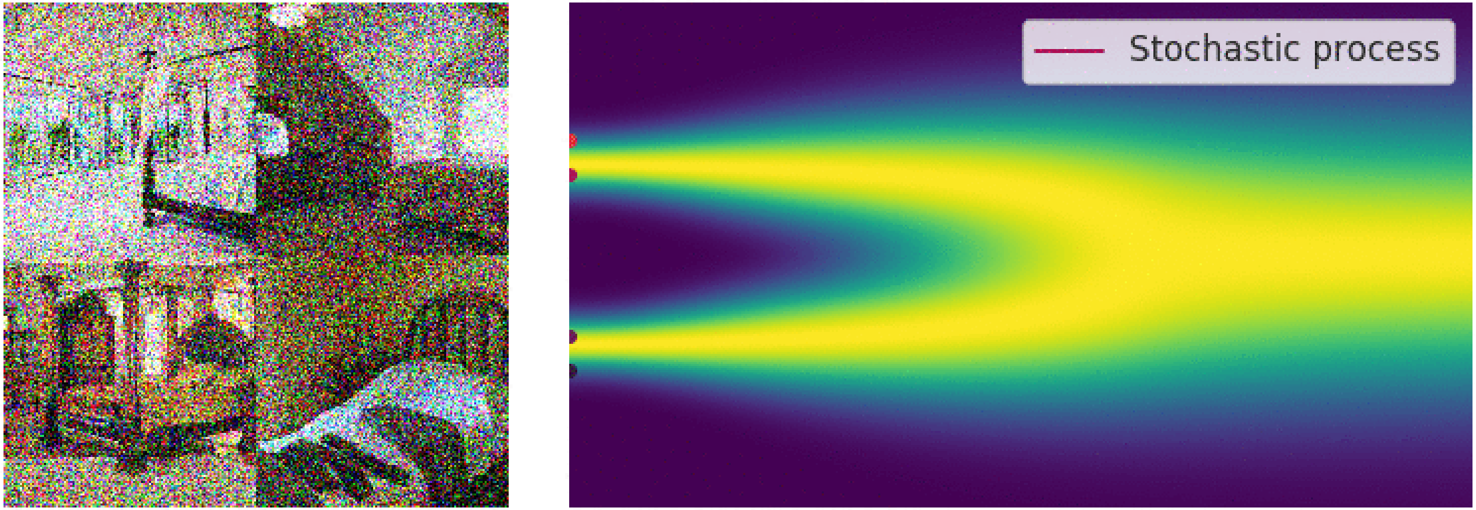
Score-based generative models 101

Samples $x \sim p(x)$ are progressively perturbed through a diffusion process described by the forward SDE

$$\mathrm{d}x_t = f_t x_t \mathrm{d}t + g_t \mathrm{d}w_t,$$

where x_t is the perturbed sample at time t, leading to a Gaussian diffusion kernel

$$p(x_t | x) = \mathcal{N}(x_t | \alpha_t x, \Sigma_t).$$

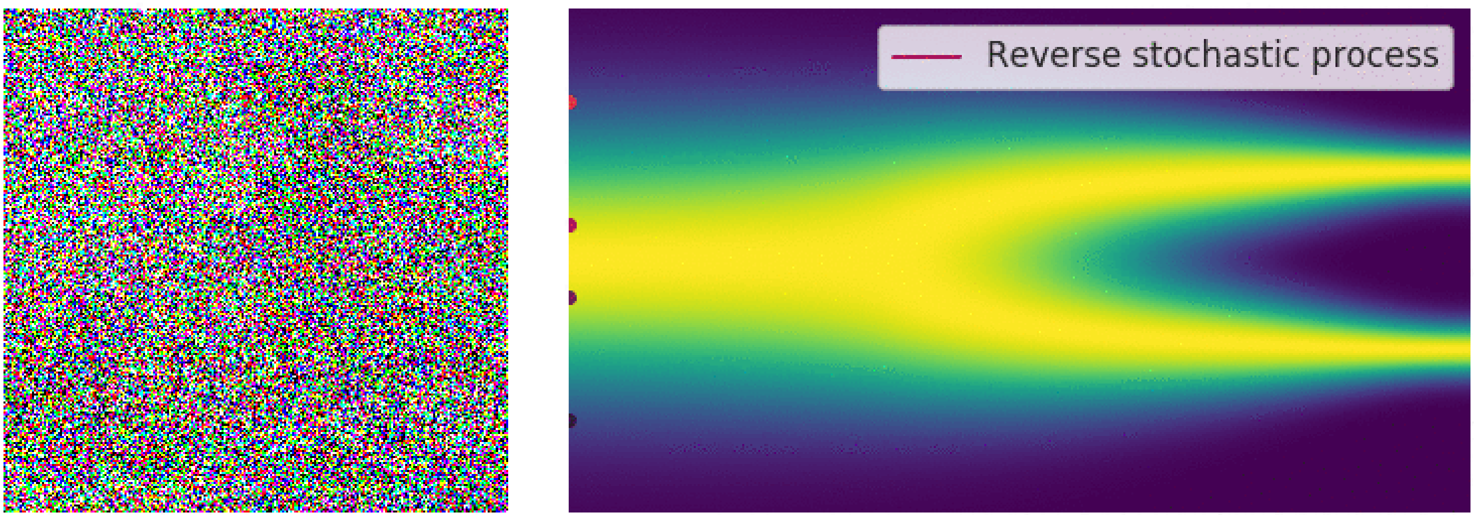


Forward diffusion process.

The reverse process satisfies a reverse-time SDE that can be derived analytically from the forward SDE as

$$\mathrm{d}x_t = \left[f_t x_t - \frac{1+\eta^2}{2} g_t^2 \nabla_{x_t} \log p(x_t) \right] \mathrm{d}t + \eta g_t \mathrm{d}w_t.$$

Therefore, to generate data samples $x_0 \sim p(x_0) \approx p(x)$, we can draw noise samples $x_1 \sim p(x_1) \approx \mathcal{N}(0, \Sigma_1)$ and gradually remove the noise therein by simulating the reverse SDE from $t = 1$ to 0.

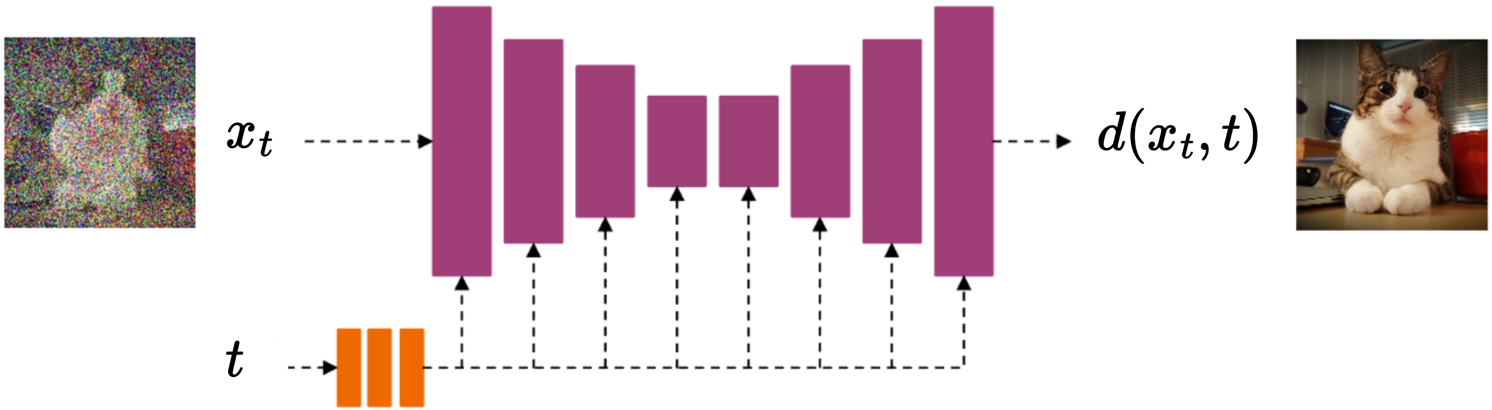


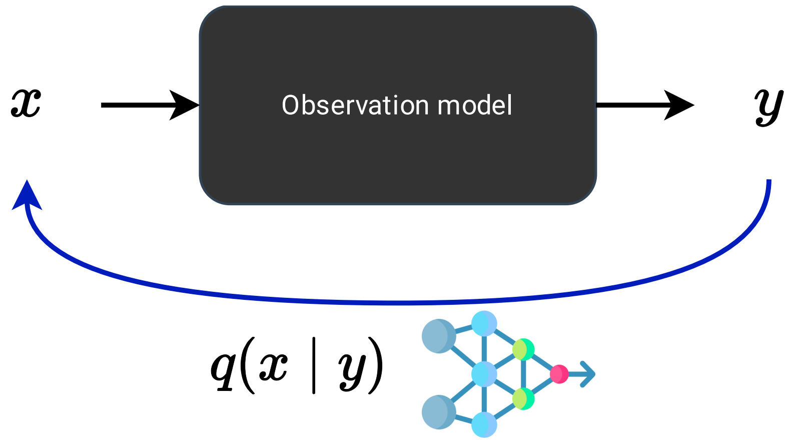
Reverse denoising process.

The score function $\nabla_{x_t} \log p(x_t)$ in the reverse SDE is unknown, but can be approximated by a neural network $d_\theta(x_t, t)$ by minimizing the denoising score matching objective

$$\mathbb{E}_{p(x)p(t)p(x_t|x)} \left[\lambda_t ||d_\theta(x_t, t) - x||_2^2\right].$$

The optimal denoiser d_θ is the mean $\mathbb{E}[x|x_t]$ which, via Tweedie's formula, allows to use $s_\theta(x_t, t) = \Sigma_t^{-1}(d_\theta(x_t, t) - x_t)$ as a score estimate in the reverse SDE.





Inverting single observations

Score-based generative models can be made conditional by extending the denoiser $d_\theta(x_t, t)$ to $d_\theta(x_t, t, y)$ to model the conditional distribution $p(x|y)$. The denoiser is then trained to minimize the conditional denoising score matching objective

$$\mathbb{E}_{p(y)p(x|y)p(t)p(x_t|x,y)} \left[\lambda_t || d_\theta(x_t, t, y) - x ||_2^2 \right] .$$

Alternatively, because of the Bayes' rule, the posterior score $\nabla_{x_t} \log p(x_t|y)$ to inject in the reverse SDE can be decomposed as

$$\nabla_{x_t} \log p(x_t|y) = \nabla_{x_t} \log p(x_t) + \nabla_{x_t} \log p(y|x_t).$$

This is particularly convenient as

- if $\nabla_{x_t} \log p(y|x_t)$ can approximated, then it enables **zero-shot posterior sampling** from a pre-trained diffusion prior $p(x_0)$, without having to pre-wire the neural denoiser to the observation model $p(y|x)$.
- it does not require paired data (x, y).

Approximating $\nabla_{x_t} \log p(y|x_t)$

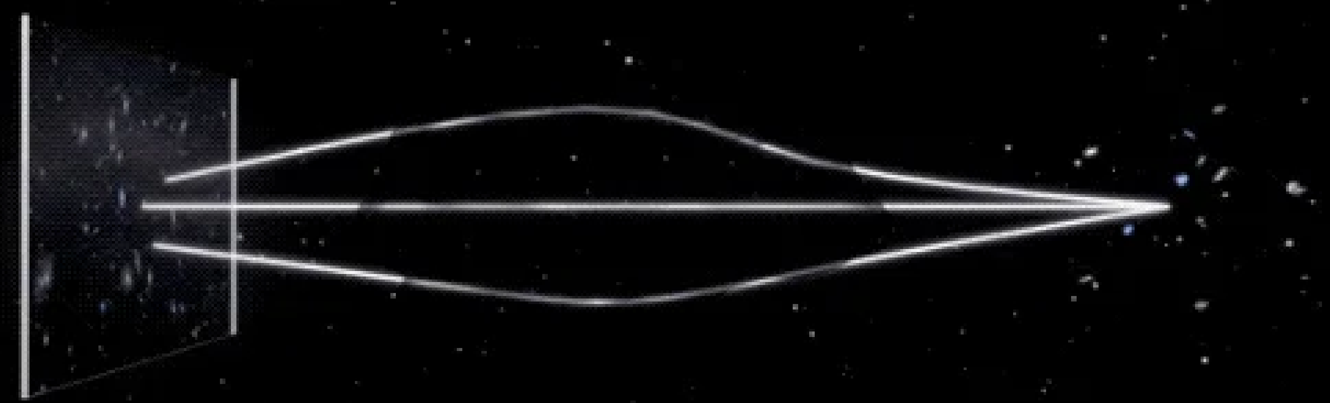
Assume a differentiable measurement function $\mathcal{A}$ and a Gaussian observation model $p(y|x) = \mathcal{N}(y|\mathcal{A}(x), \Sigma_y)$.

We want to estimate the score $\nabla x_t \log p(y|x_t)$ of the noise-perturbed likelihood

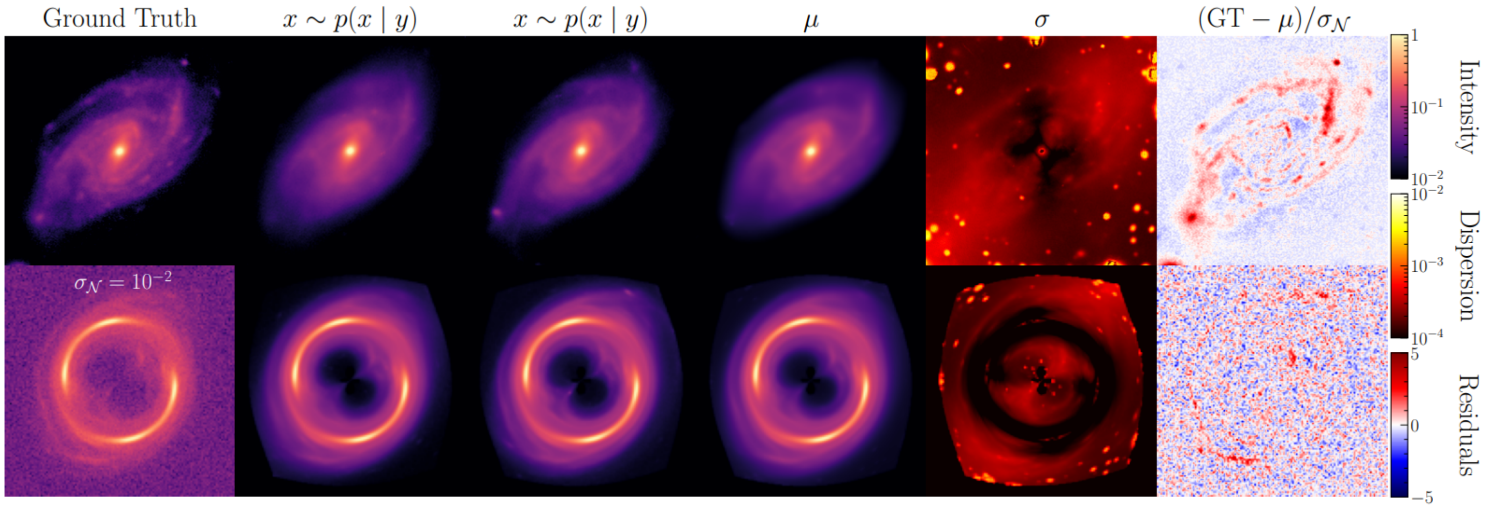
$$p(y|x_t) = \int p(y|x)p(x|x_t)\mathrm{d}x.$$

- DPS (Chung et al, 2022): $p(y|x_t) \approx \mathcal{N}(y|\mathcal{A}(\mathbb{E}[x|x_t]), \Sigma_y)$.
- SDA (Rozet and Louppe, 2023): $p(y|x_t) \approx \mathcal{N}(y|\mathcal{A}(\mathbb{E}[x|x_t]), \Sigma_y + \frac{\sigma_t^2}{\mu_t^2} A\Gamma A^T)$
- MMPS (Rozet et al, 2024, for $\mathcal{A}(x) = Ax$): $p(y|x_t) \approx \mathcal{N}(y|A\mathbb{E}[x|x_t], \Sigma_y + A\mathbb{V}[x|x_t]A^T)$, where $\mathbb{V}[x|x_t]$ is estimated using Tweedie's covariance formula and the conjugate gradient method.

Since all these approximations are Gaussian, the score of the noise-perturbed likelihood can be estimated analytically from their parameterization.



Example: Inverting gravitational lensing observations.

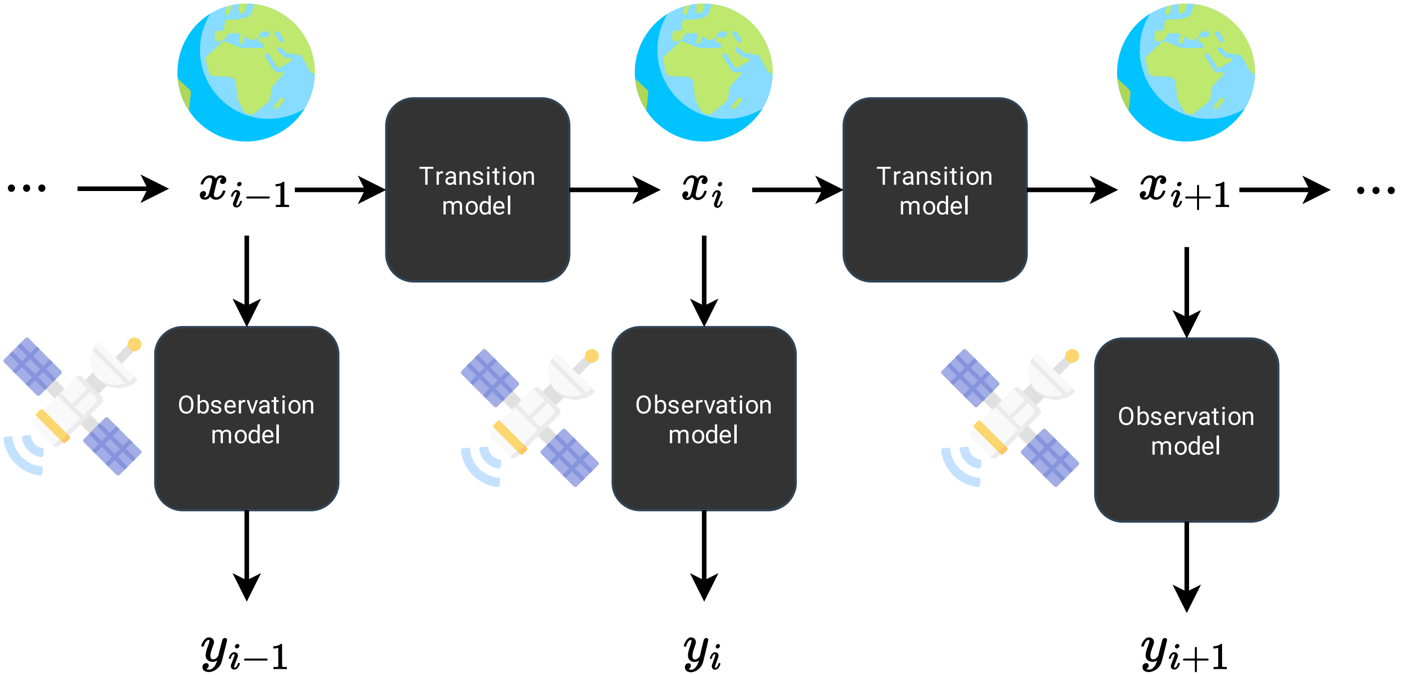


If $p(x)$ is a diffusion prior over regular galaxy images and $p(y|x)$ is a lensing operator, observations y can be inverted to recover the original galaxy images x by posterior sampling from $p(x|y)$.

—

Credits: Adam et al., 2022 (arXiv:2211.03812).

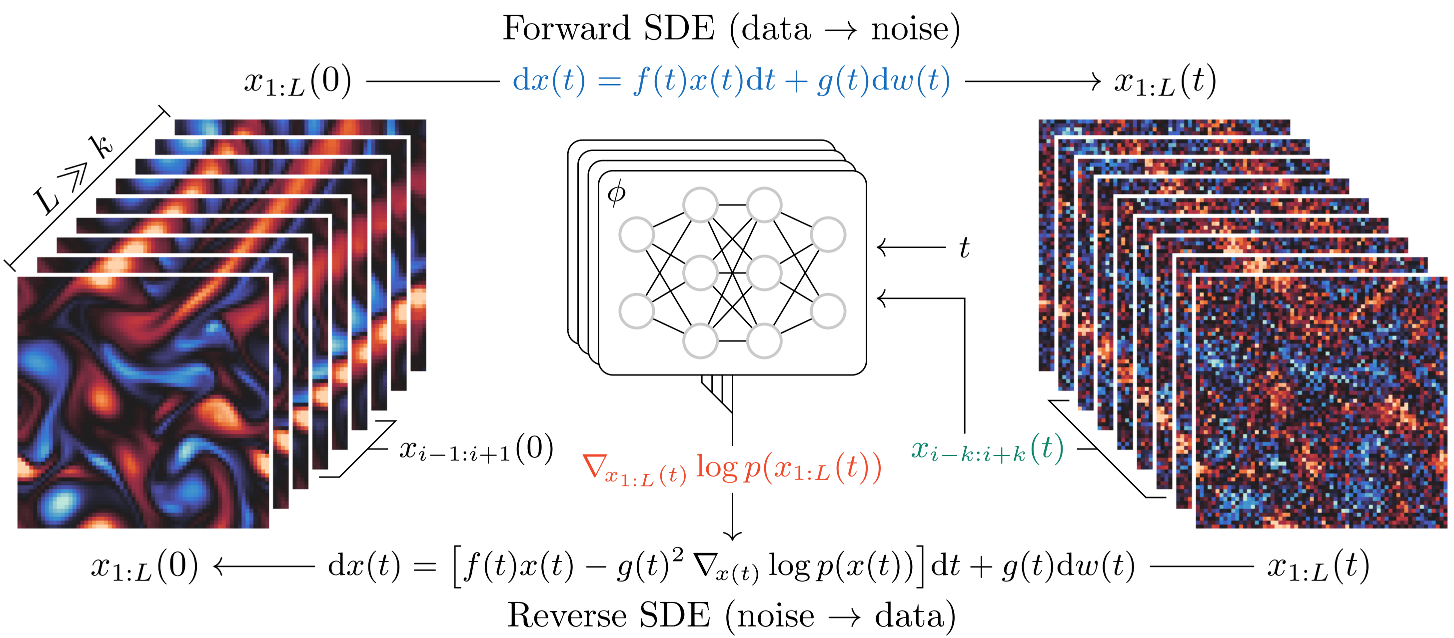
Score-based data assimilation in dynamical systems



Assume the latent state x evolves according to a transition model $p(x_{i+1} | x_i)$ and is observed through an observation model $p(y | x_{1:L})$. (Typically, the observation model will be $p(y_i | x_i)$, but we consider the general case here.)

The goal of **data assimilation** is to estimate plausible trajectories $x_{1:L}$ given one or more noisy observations y (or $y_{1:L}$), that is to estimate the posterior

$$p(x_{1:L} | y) = \frac{p(y | x_{1:L})}{p(y)} p(x_0) \prod_{i=1}^{L-1} p(x_{i+1} | x_i).$$

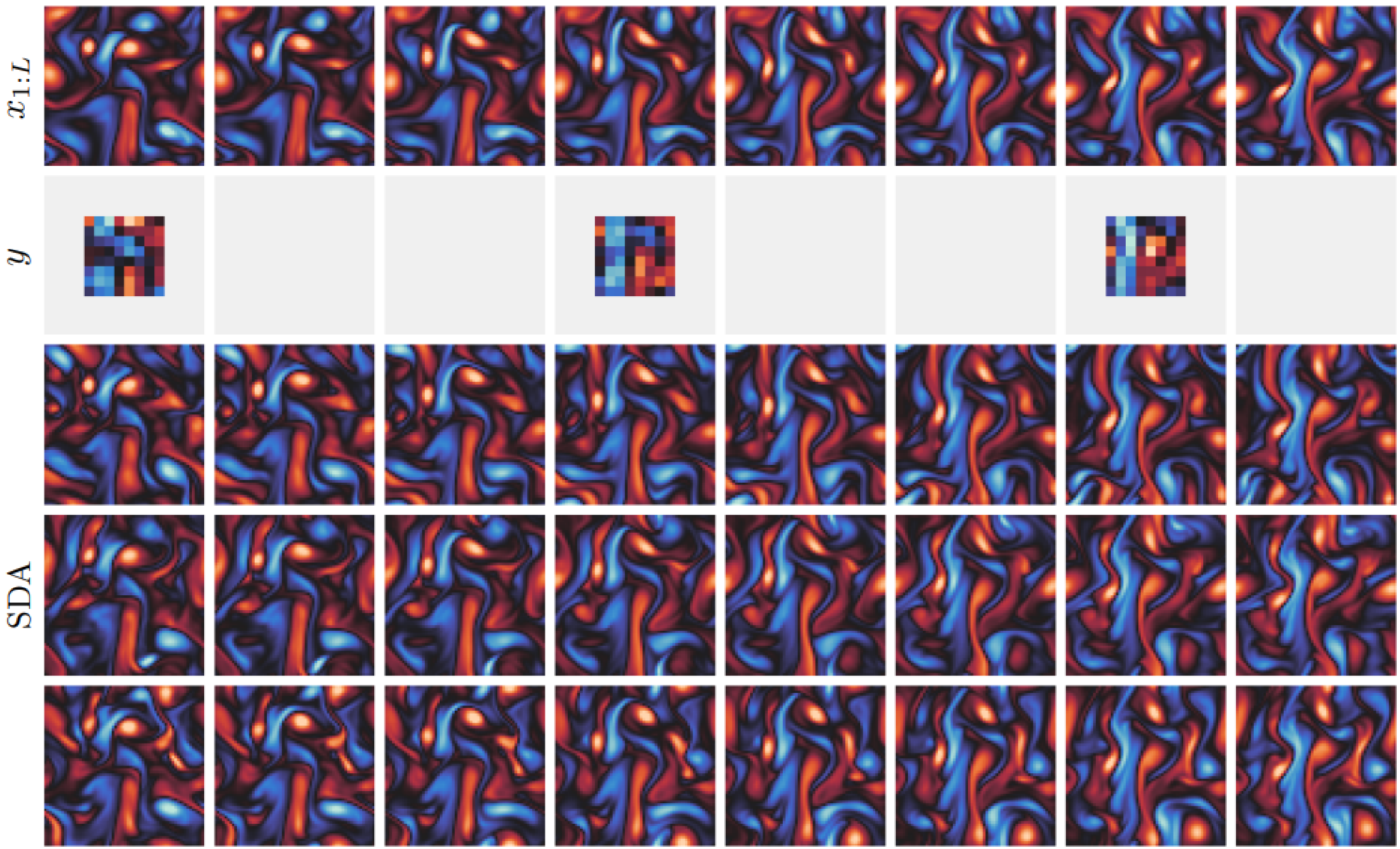


Our approach:

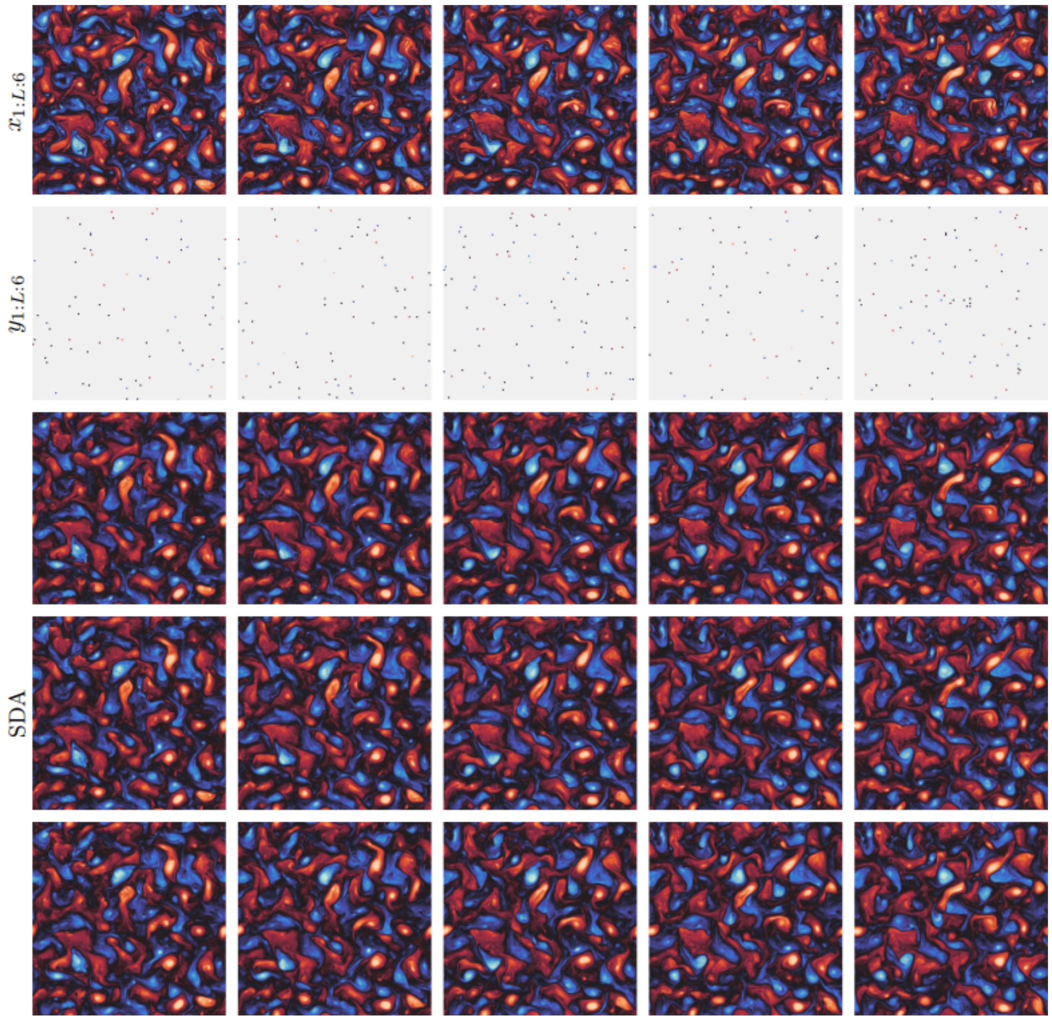
- Build a score-based generative model $p(x_{1:L})$ of arbitrary-length trajectories*.
- Use zero-shot posterior sampling to generate plausible trajectories from noisy observations y.

—

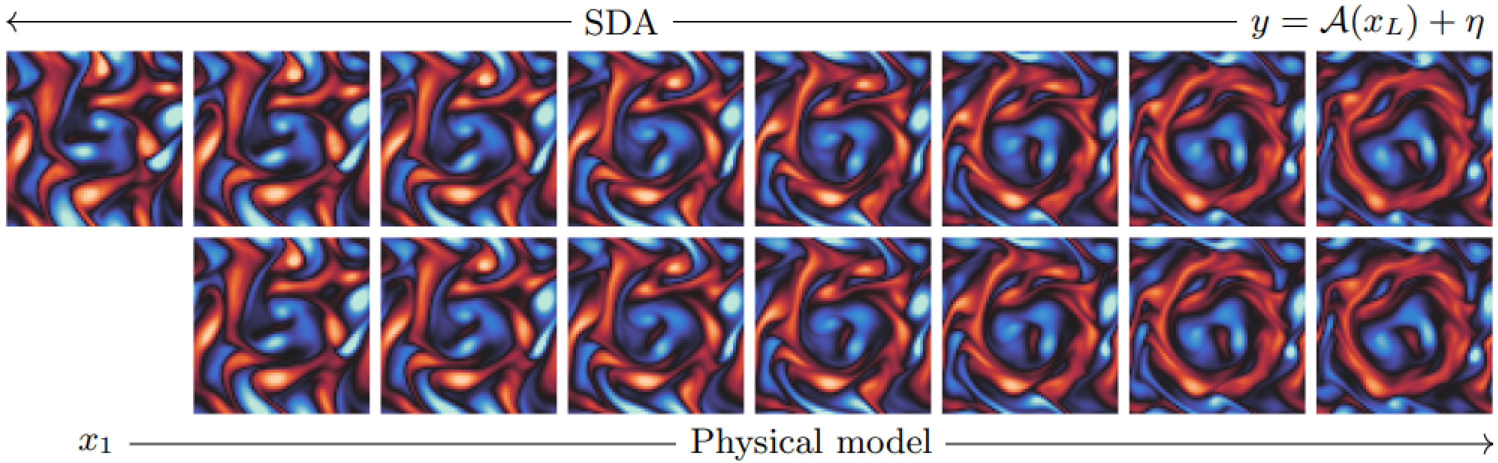
*: The score of a (noise perturbed) trajectory can be approximated by a sum of scores. See paper for details.



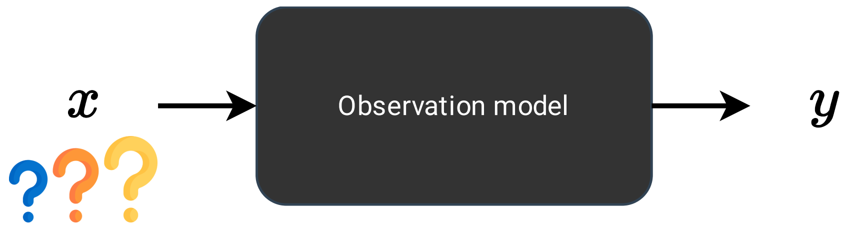
Sampling trajectories from noisy, incomplete and coarse-grained observations.



Sampling trajectories of $256 \times 256 \times 6$-dimensional states from a two-layer quasi-geostrophic model.



Sampling physically-consistent trajectories from implausible constraints.



Learning priors from noisy observations

Assume only observations $y \sim p(y)$ and a known observation model $p(y|x)$.

The objective of **Empirical Bayes** is find a prior model $q_\theta(x)$ such that

$$q_\theta(y) = \int p(y|x) q_\theta(x) \mathrm{d}x$$

is closest to $p(y)$.

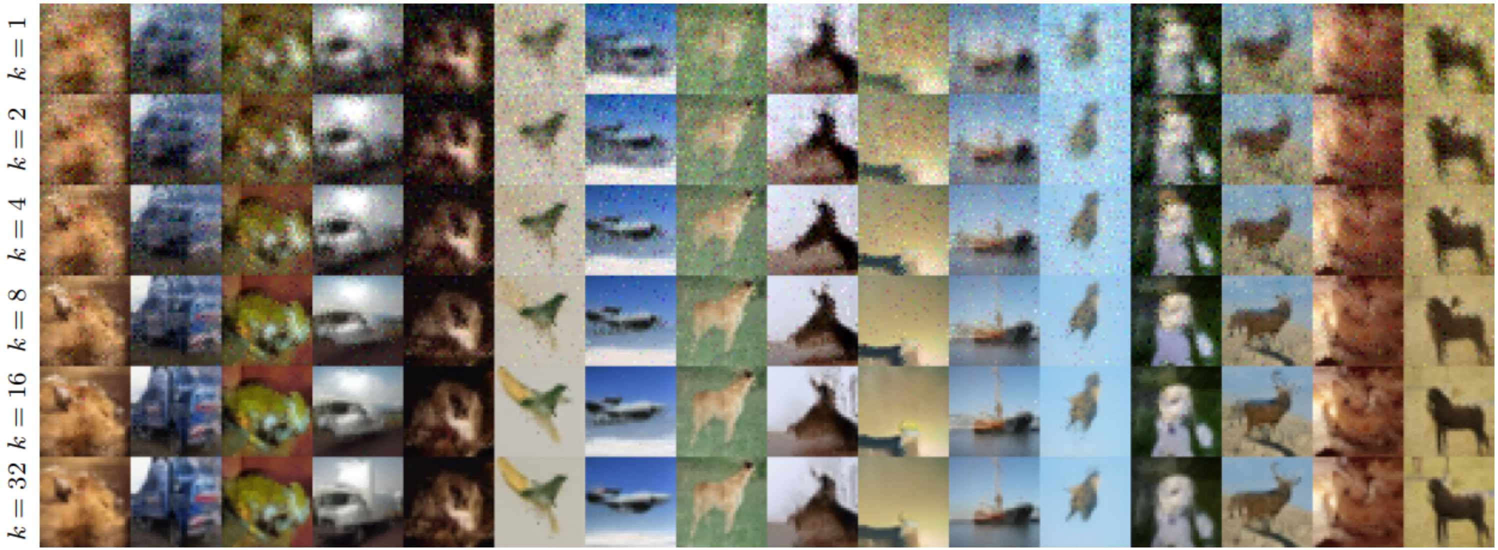
Our approach:

- If we parameterize the latent state x with a diffusion prior $q_\theta(x)$, then Expectation-Maximization can be used to maximize $q_\theta(y)$.

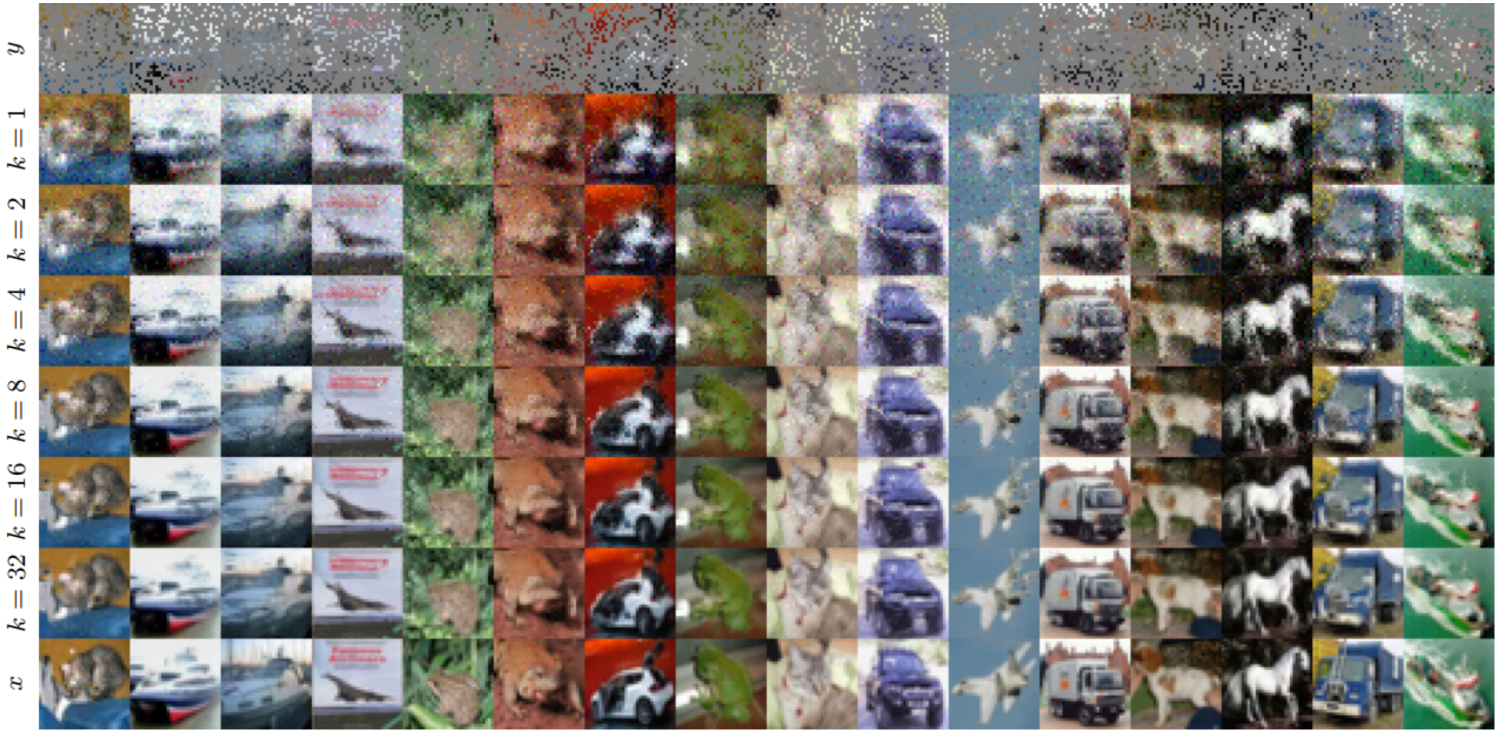
- It can be shown that the EM update

$$\theta_{k+1} = \arg\max_\theta \mathbb{E}_{p(y)} \mathbb{E}_{q_{\theta_k}(x|y)} \left[\log q_\theta(x) \right],$$

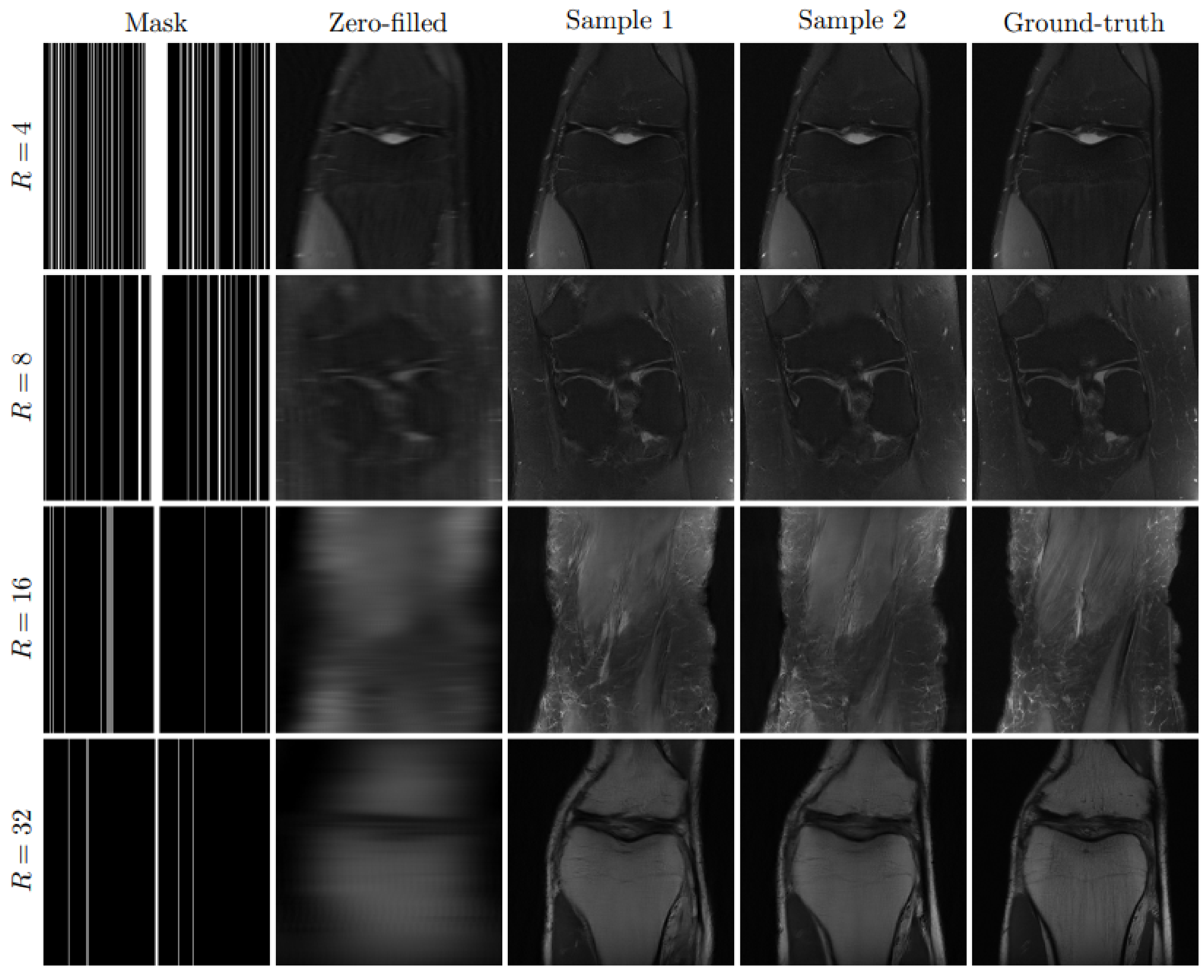
where $q_{\theta_k}(x|y)$ is obtained by posterior sampling from $q_{\theta_k}(x)$, leads to a sequence of parameters θ_k such that $\mathbb{E}_{p(y)} \left[\log q_{\theta_k}(y) \right]$ is monotonically increasing and converges to a local optimum.



Samples from the prior $q_{\theta_k}(x)$ along the EM iterations when training from corrupted CIFAR-10 images.



Samples from the posterior $q_{\theta_k}(x|y)$ along the EM iterations when training from corrupted CIFAR-10 images.



Posterior samples for accelerated MRI using a diffusion prior trained only from observations with subsampled frequencies.

Conclusions

Score-based generative models...

- can be used for high-dimensional inverse problems;
- enable zero-shot posterior sampling, without pre-wiring the network to observations;
- do not require paired data.

Next challenges:

- Rigorous diagnostics for the quality of the approximation;
- Scalability to even larger dimensions (Earth-scale weather models, videos);

The end.