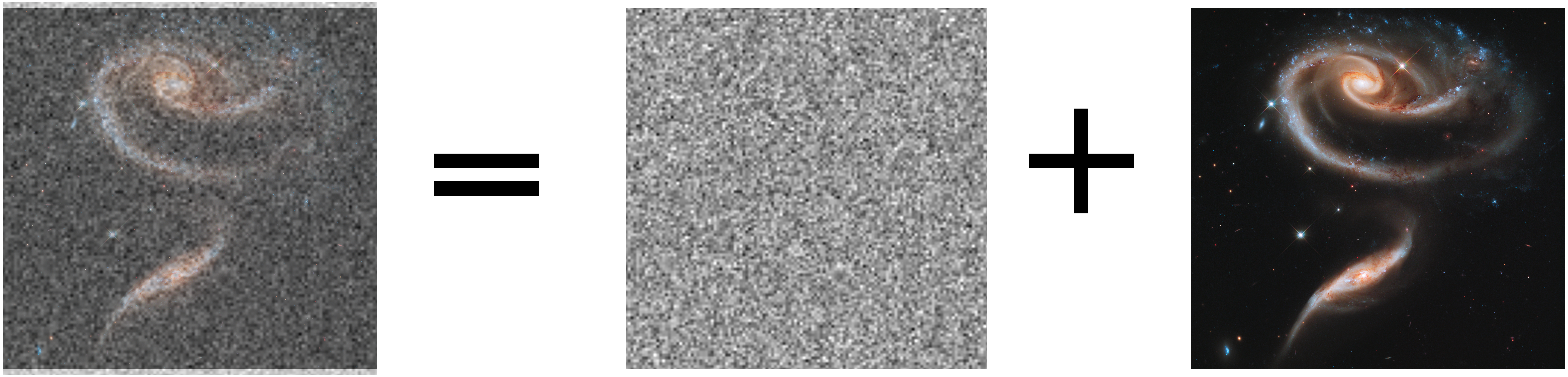
Overcoming inference challenges using score generative models

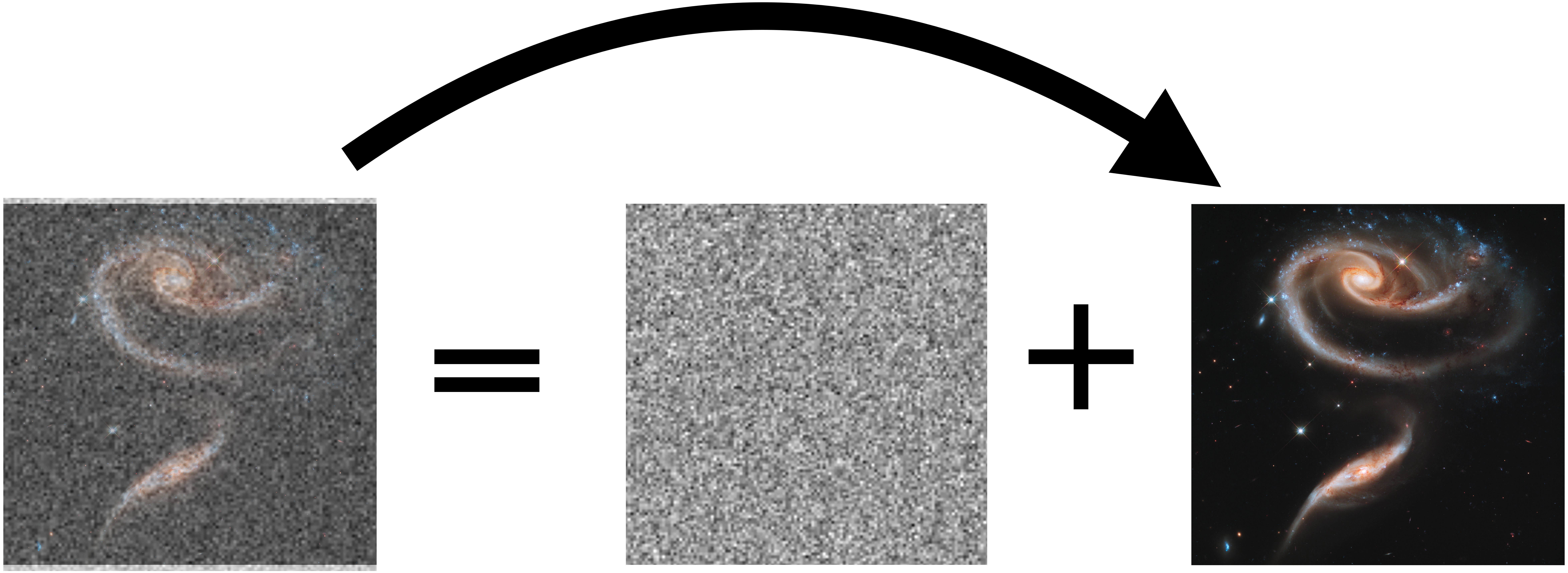
Ronan Legin

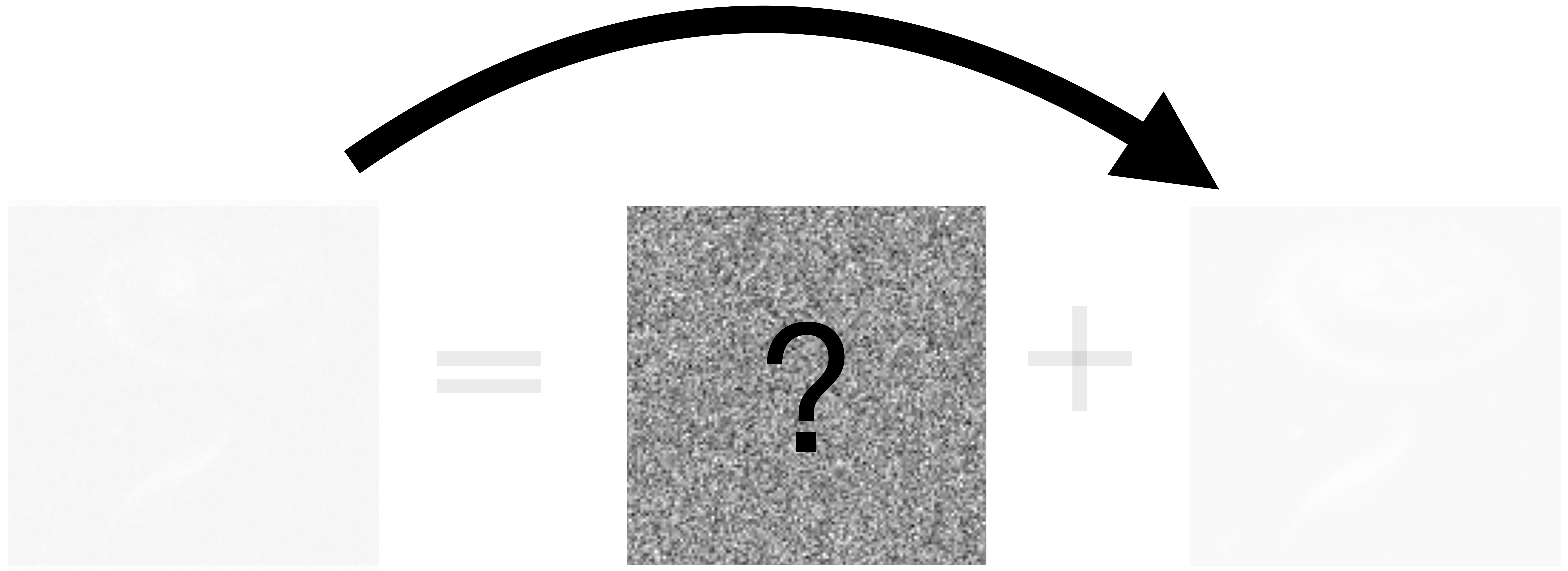
Ciela Institute

Mila

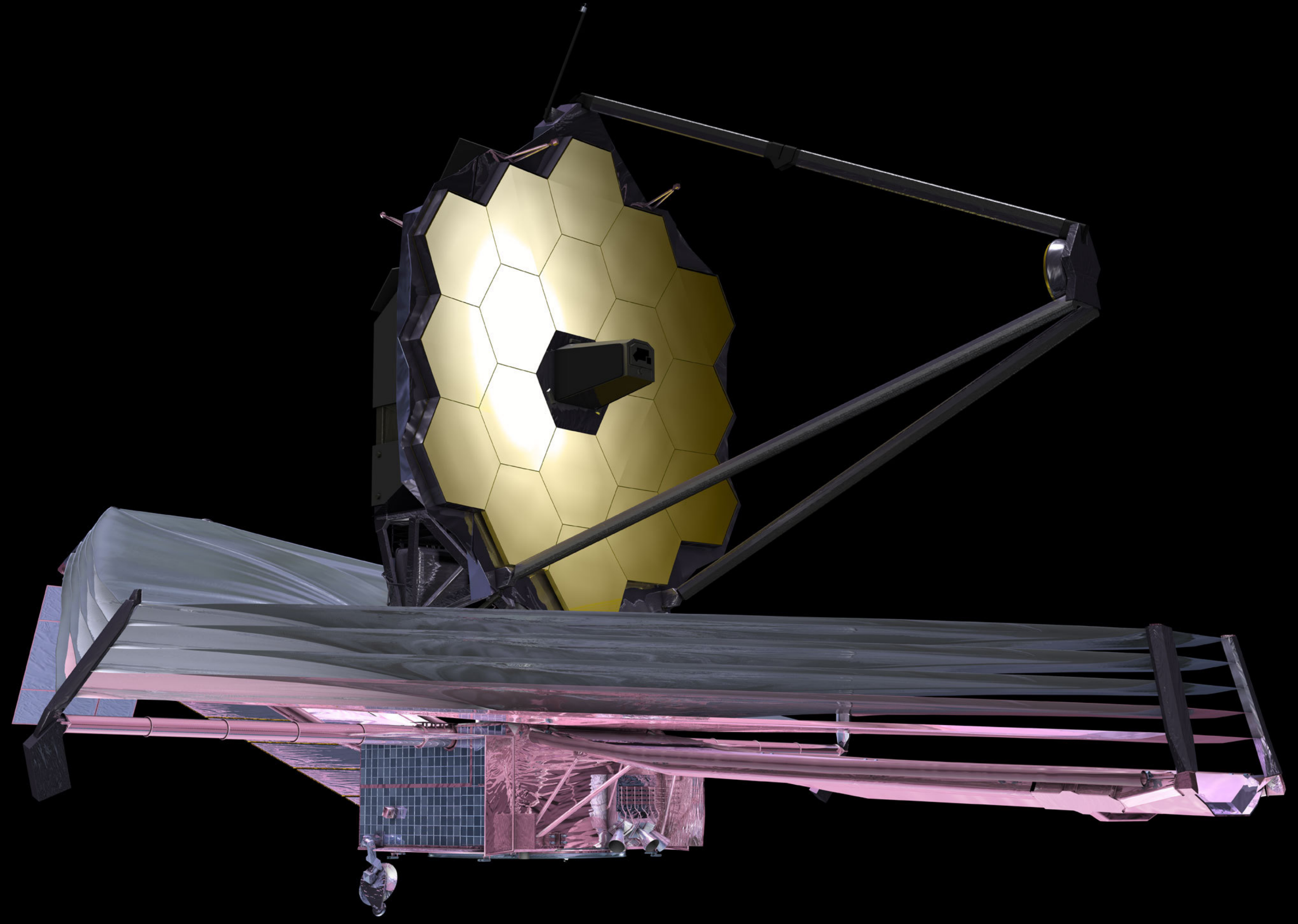
Université de Montréal



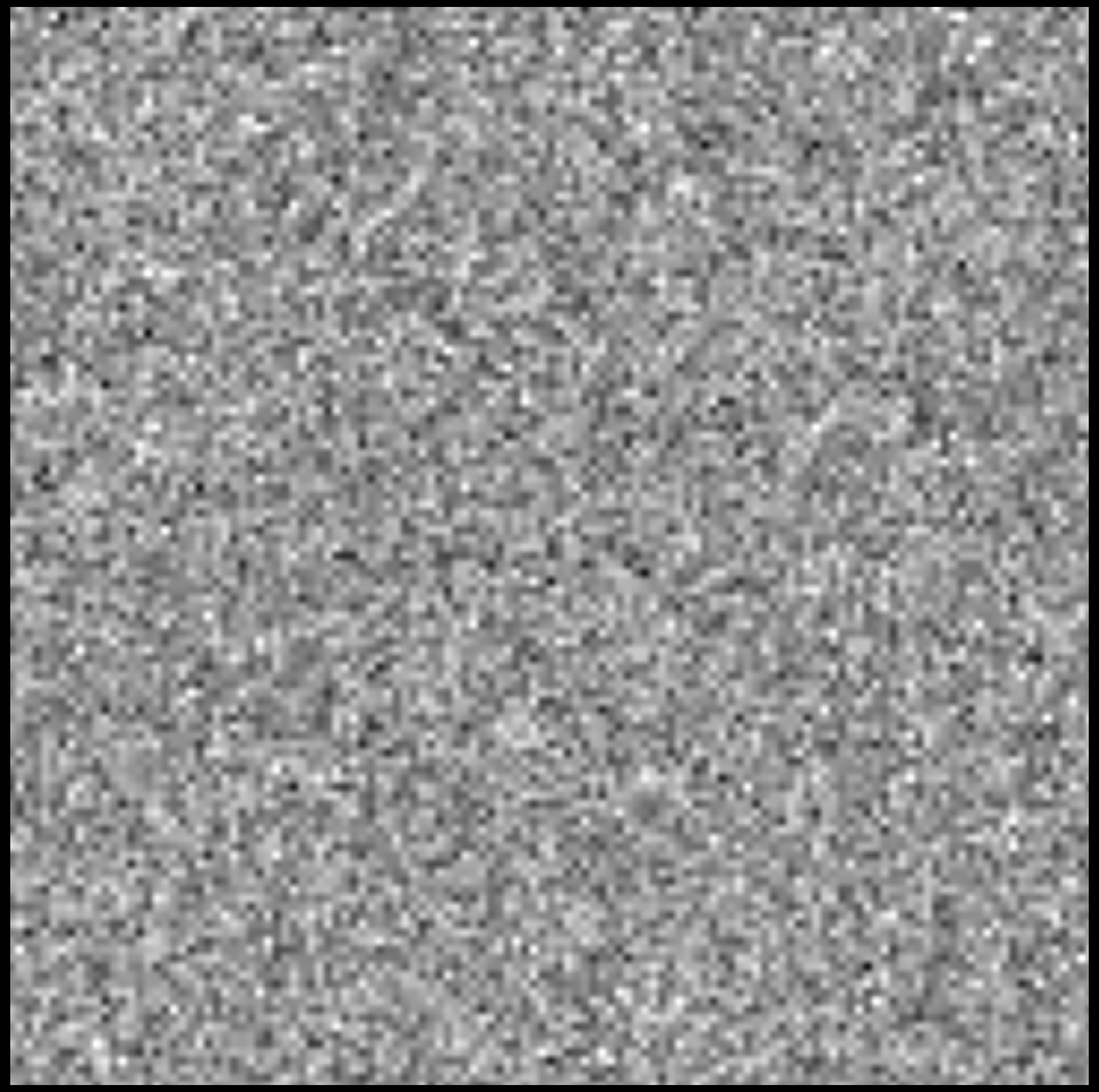




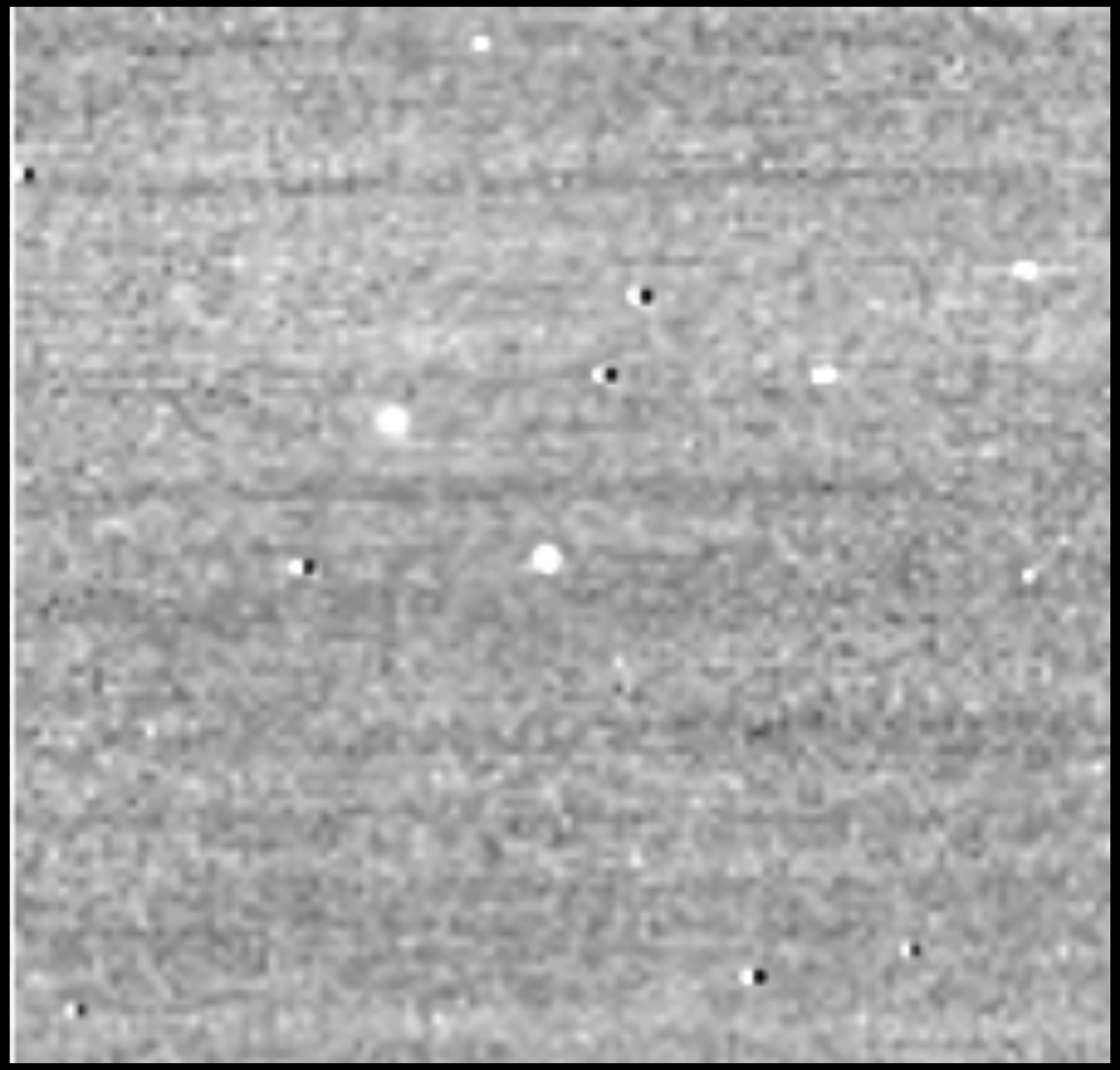
Examples

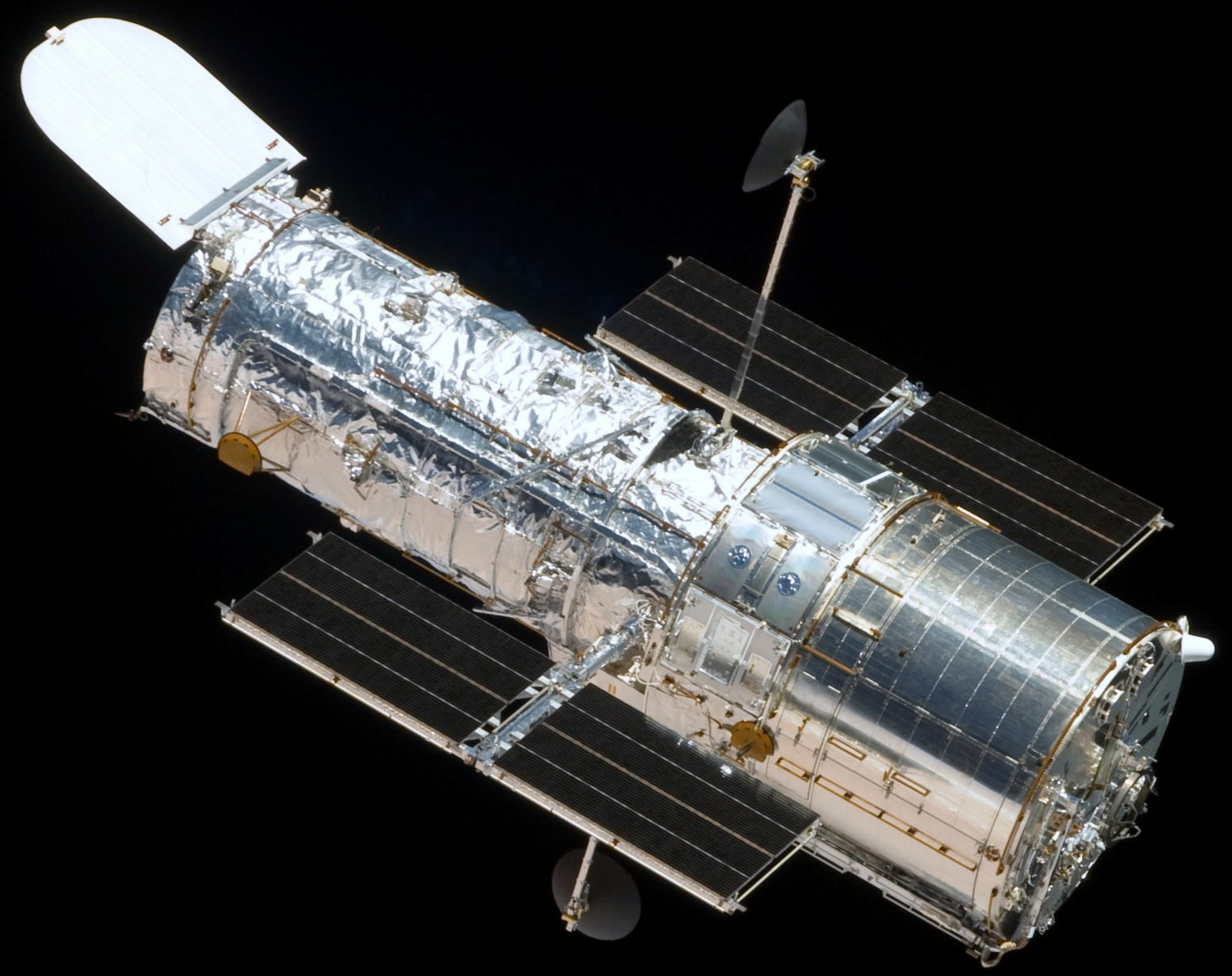


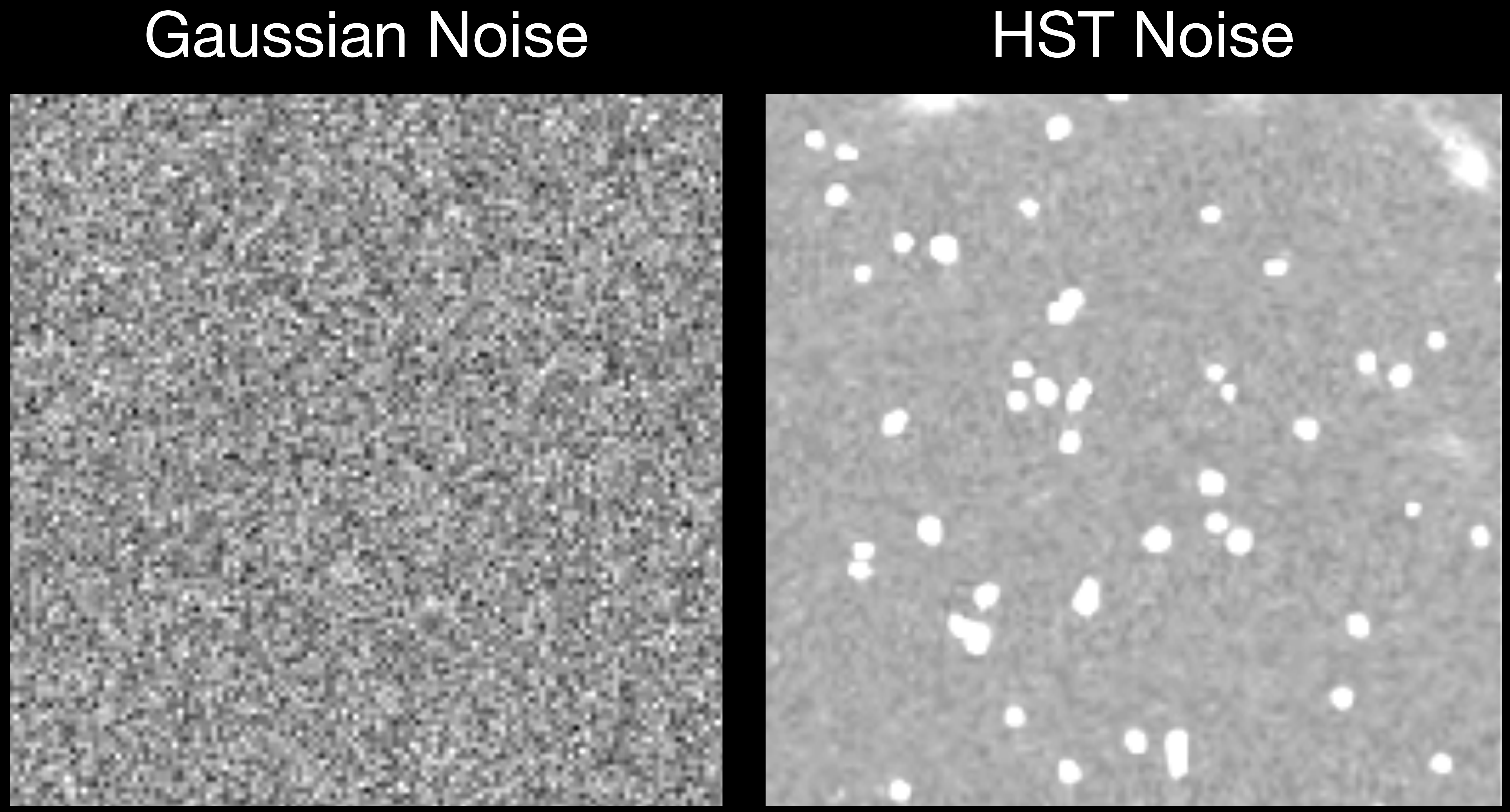
Gaussian Noise



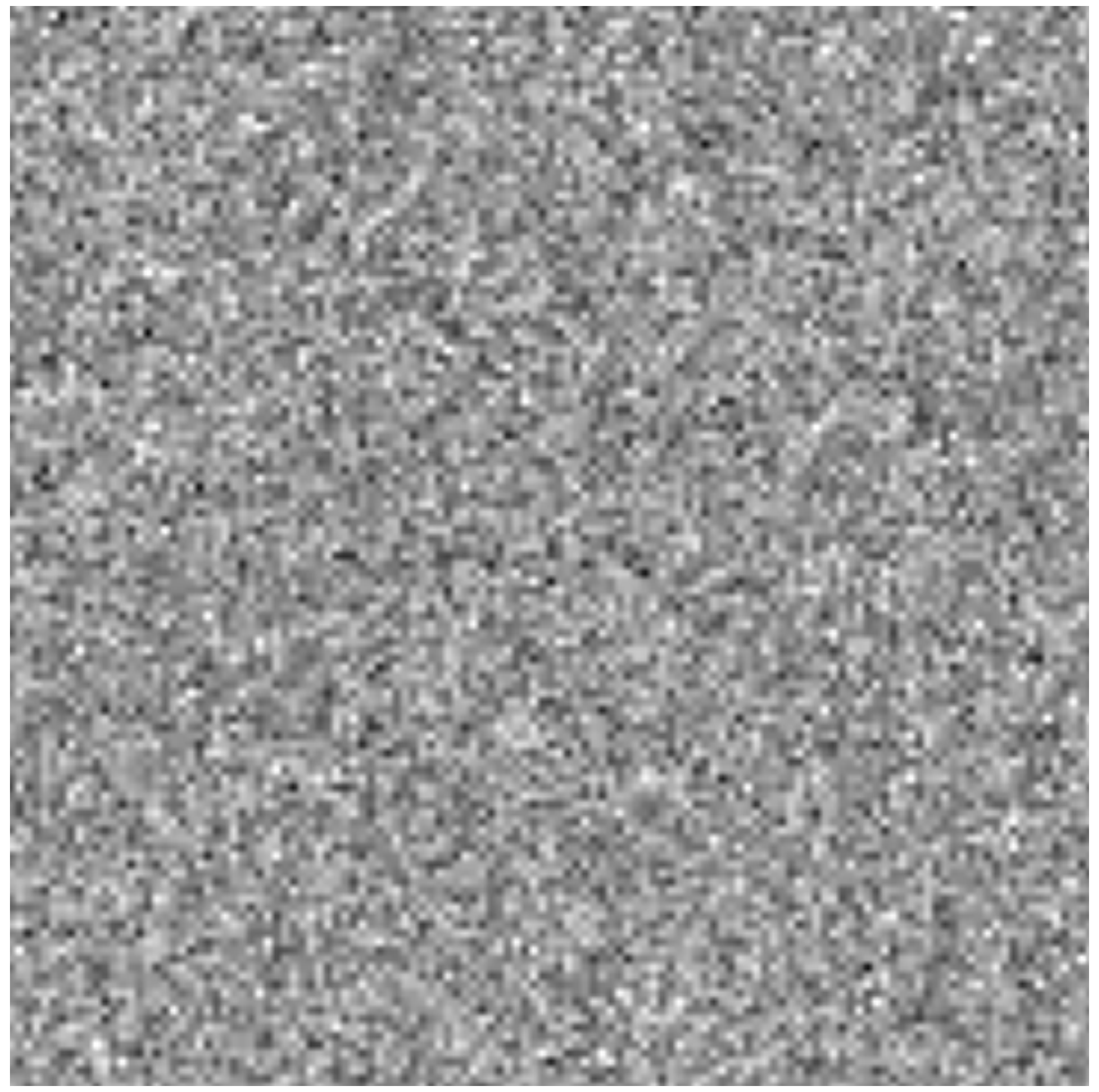
JWST Noise





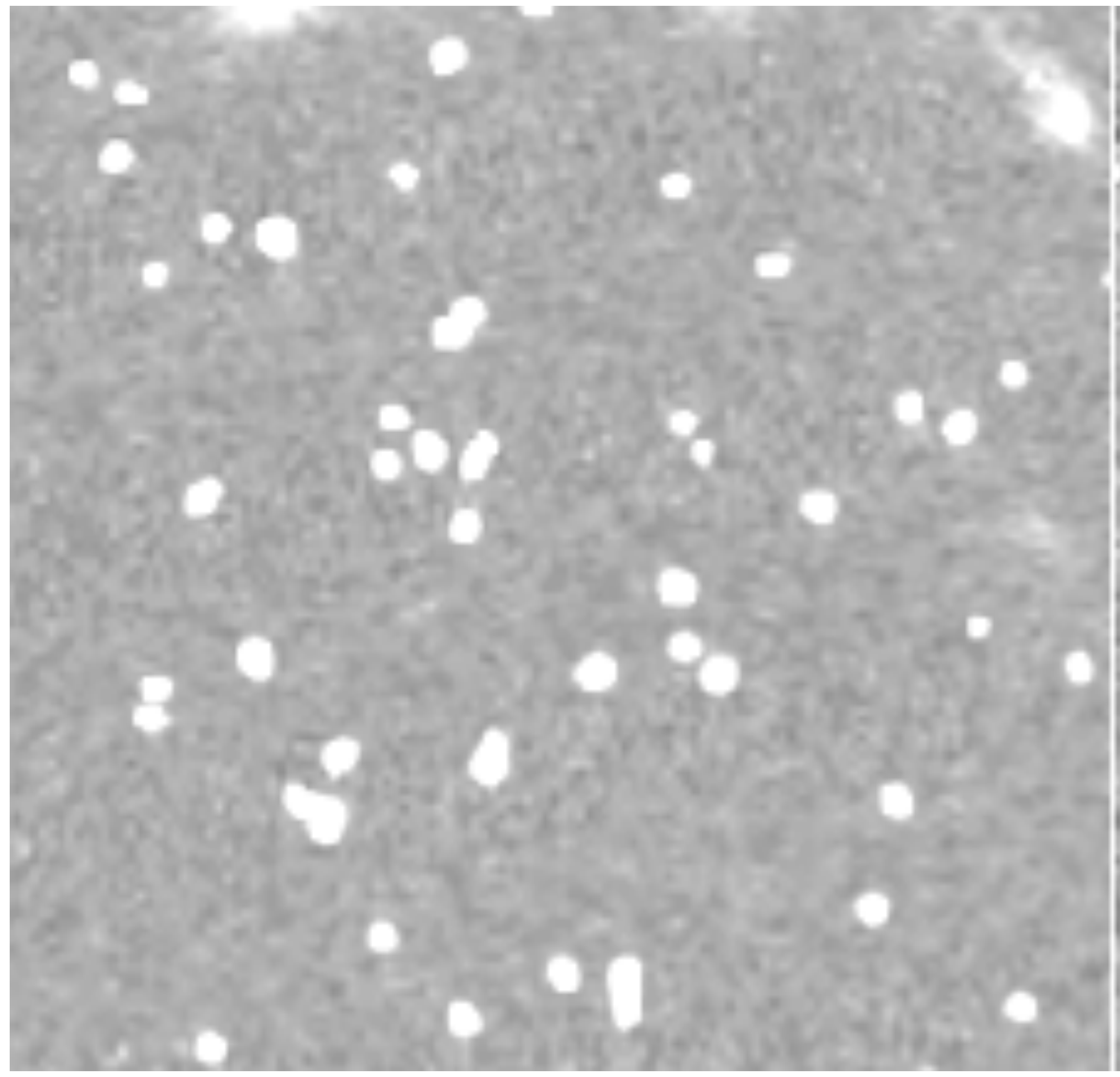

Gaussian Noise
HST Noise

Gaussian Noise



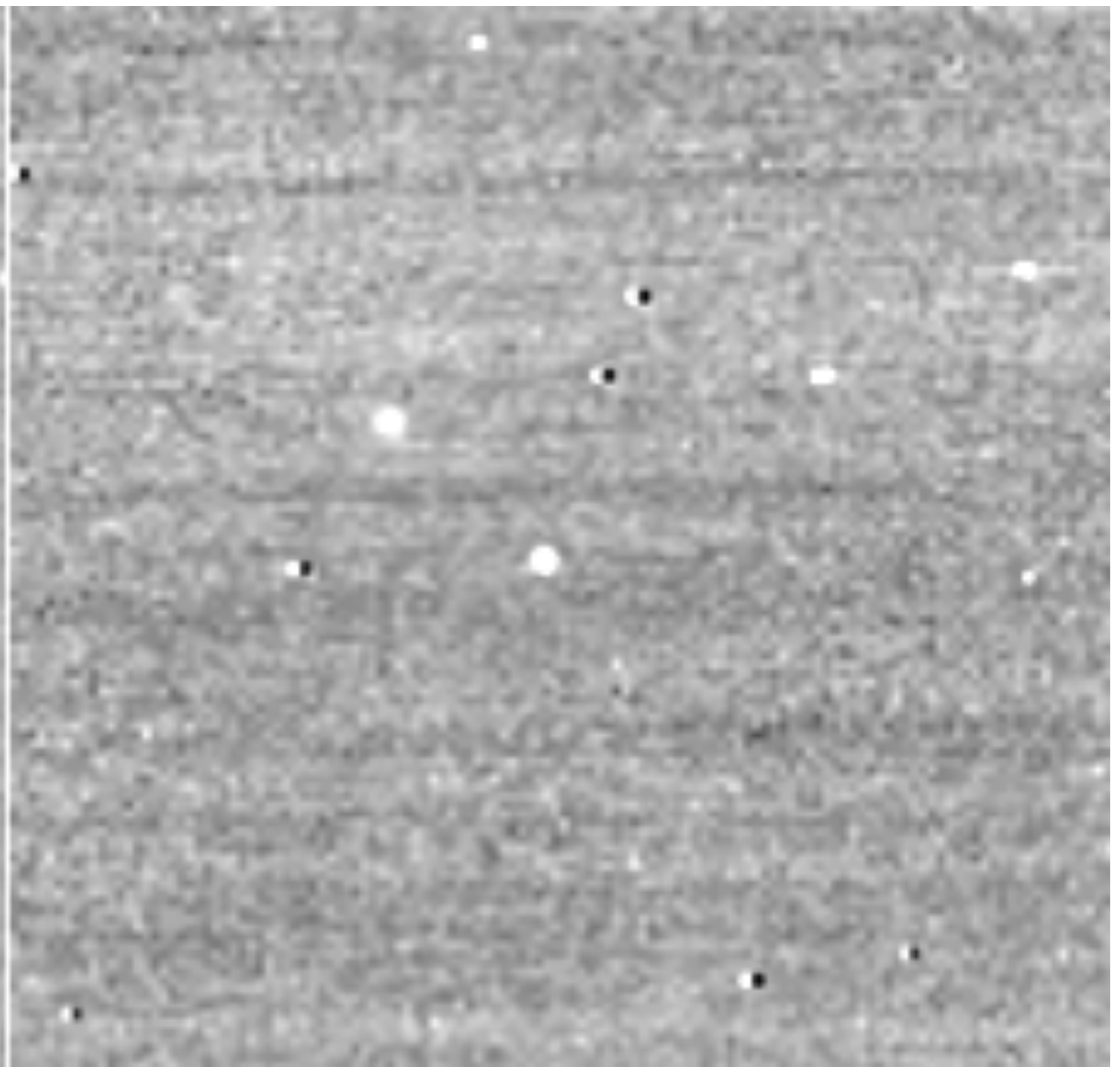
$$P(d\,|\,\theta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp - \frac{1}{2}\left(\frac{d-\mu(\theta)}{\sigma}\right)^2$$

HST noise



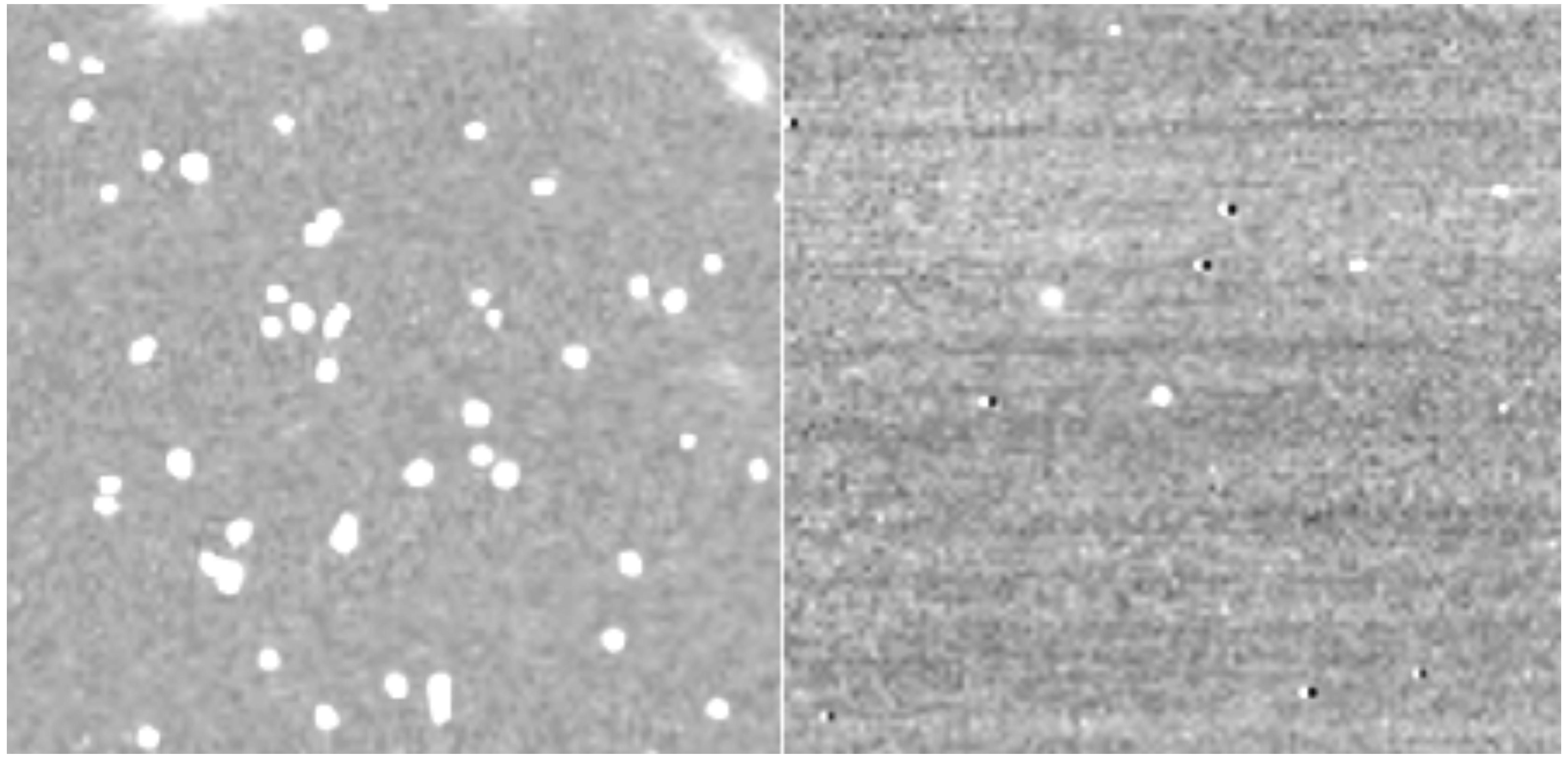
?

JWST noise

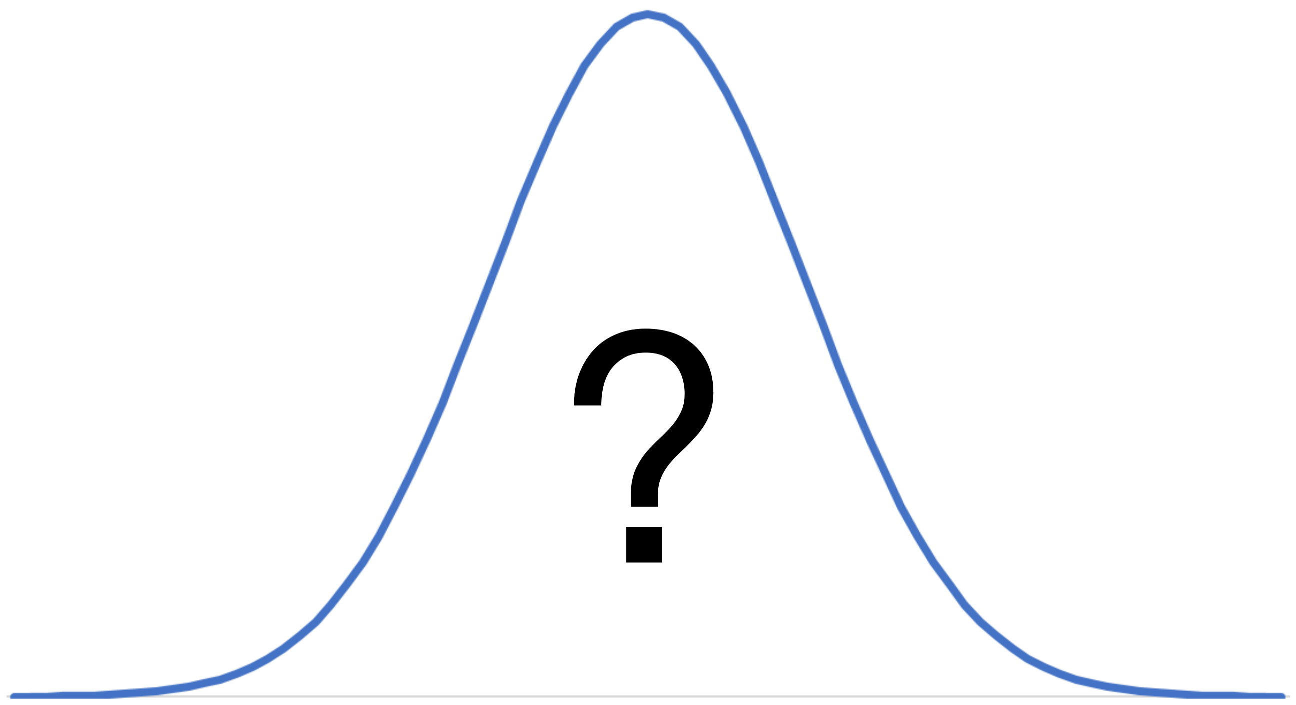


?

Learn Noise Distribution



$$Q(x)$$

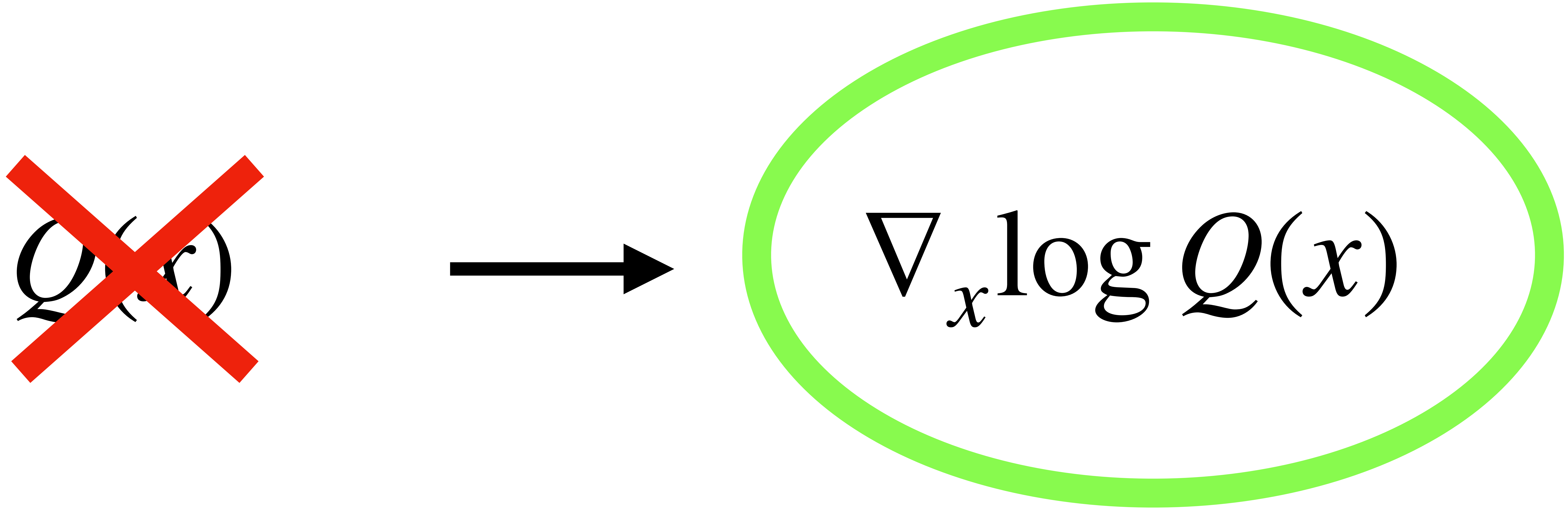


Alternative?

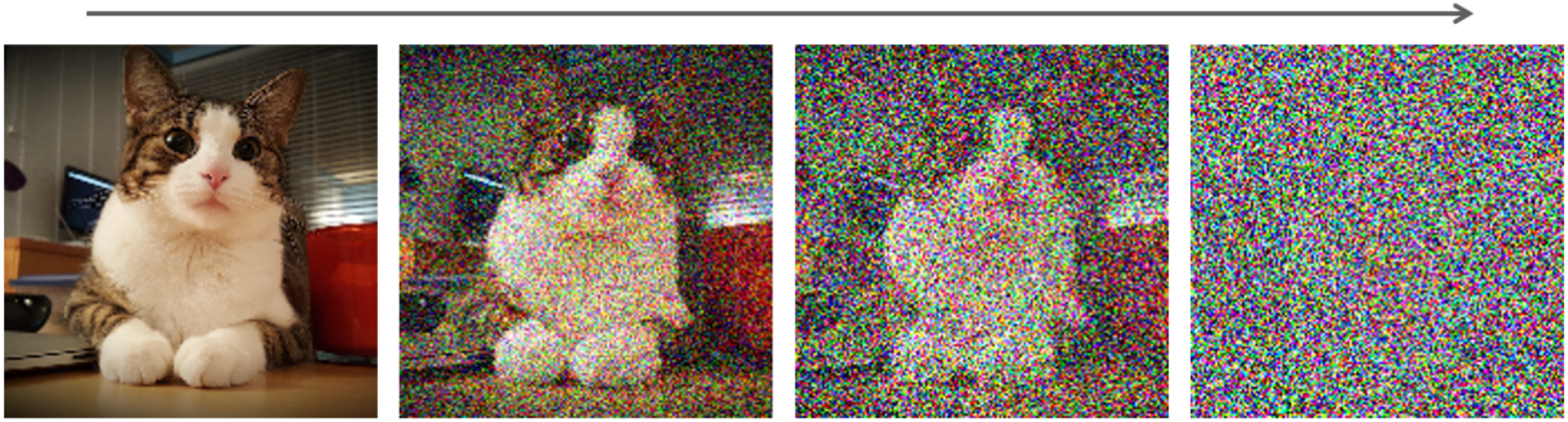
$$Q(x)$$

Alternative?

- Instead learn $\nabla_x \log Q(x)$.

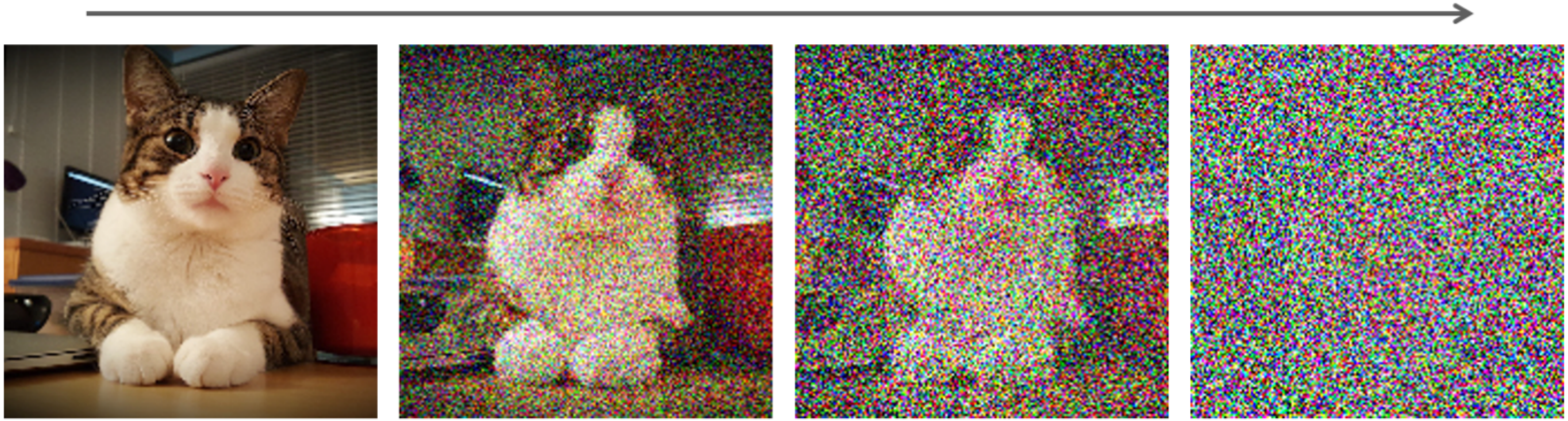


Score-based generative models



$$\nabla_x \log Q(x)$$

Score-based generative models

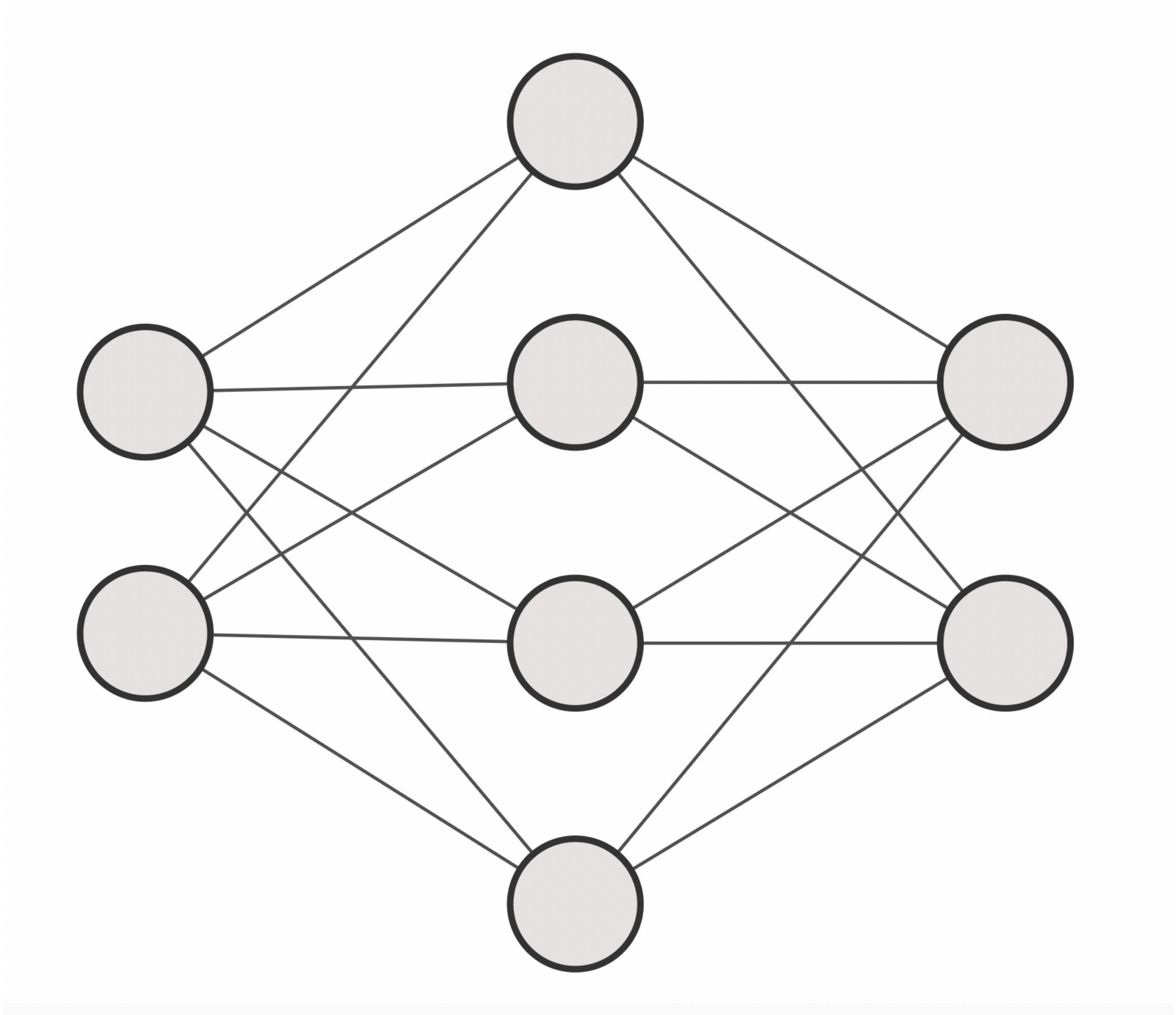


$$\nabla_x \log Q(x)$$

The Machine Learning

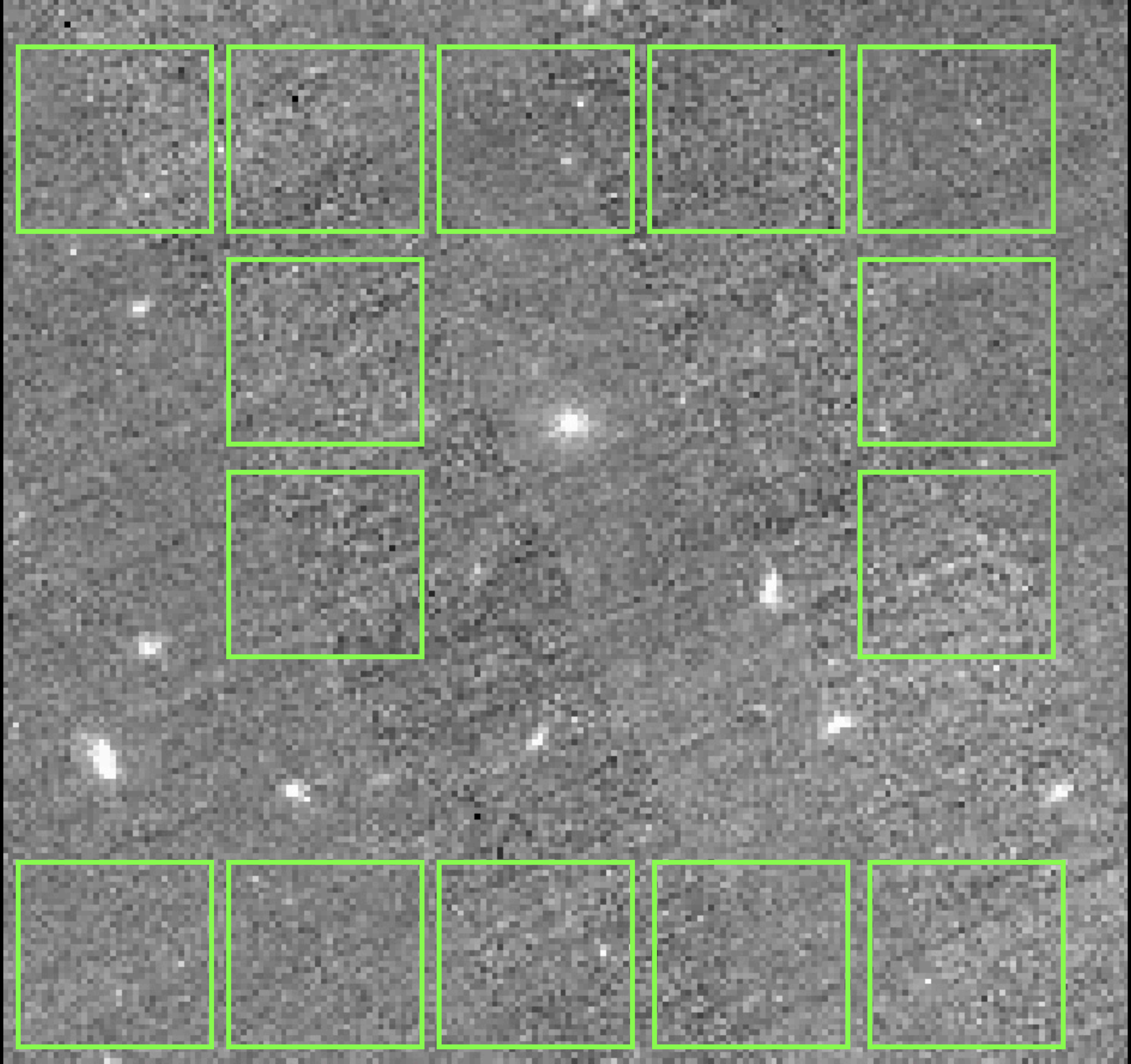
- Train neural network to predict $\nabla_x \log Q(x)$ using score matching techniques.

$$x \longrightarrow$$



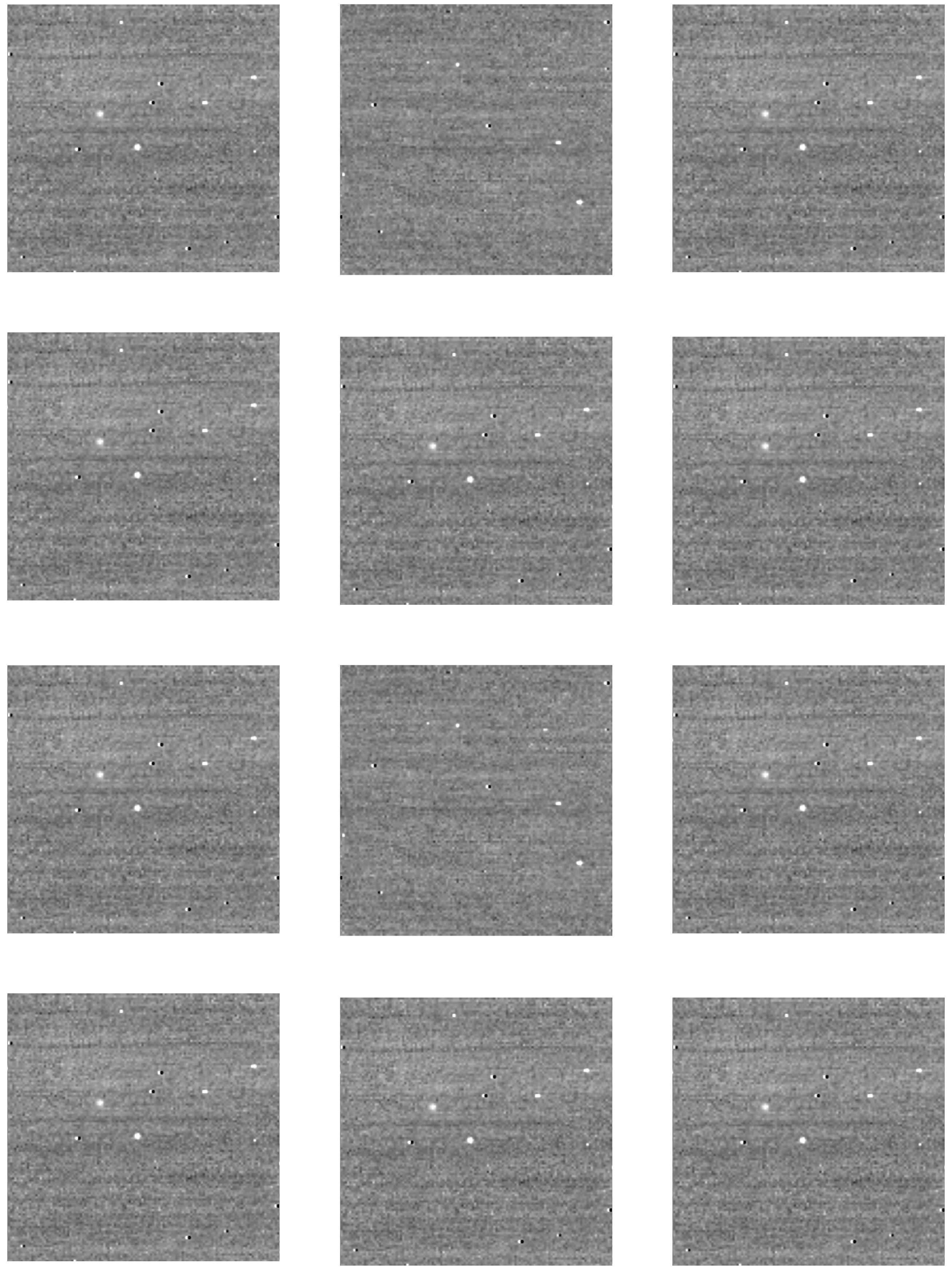
$$\longrightarrow \nabla_x \log Q(x)$$

Machine Learning

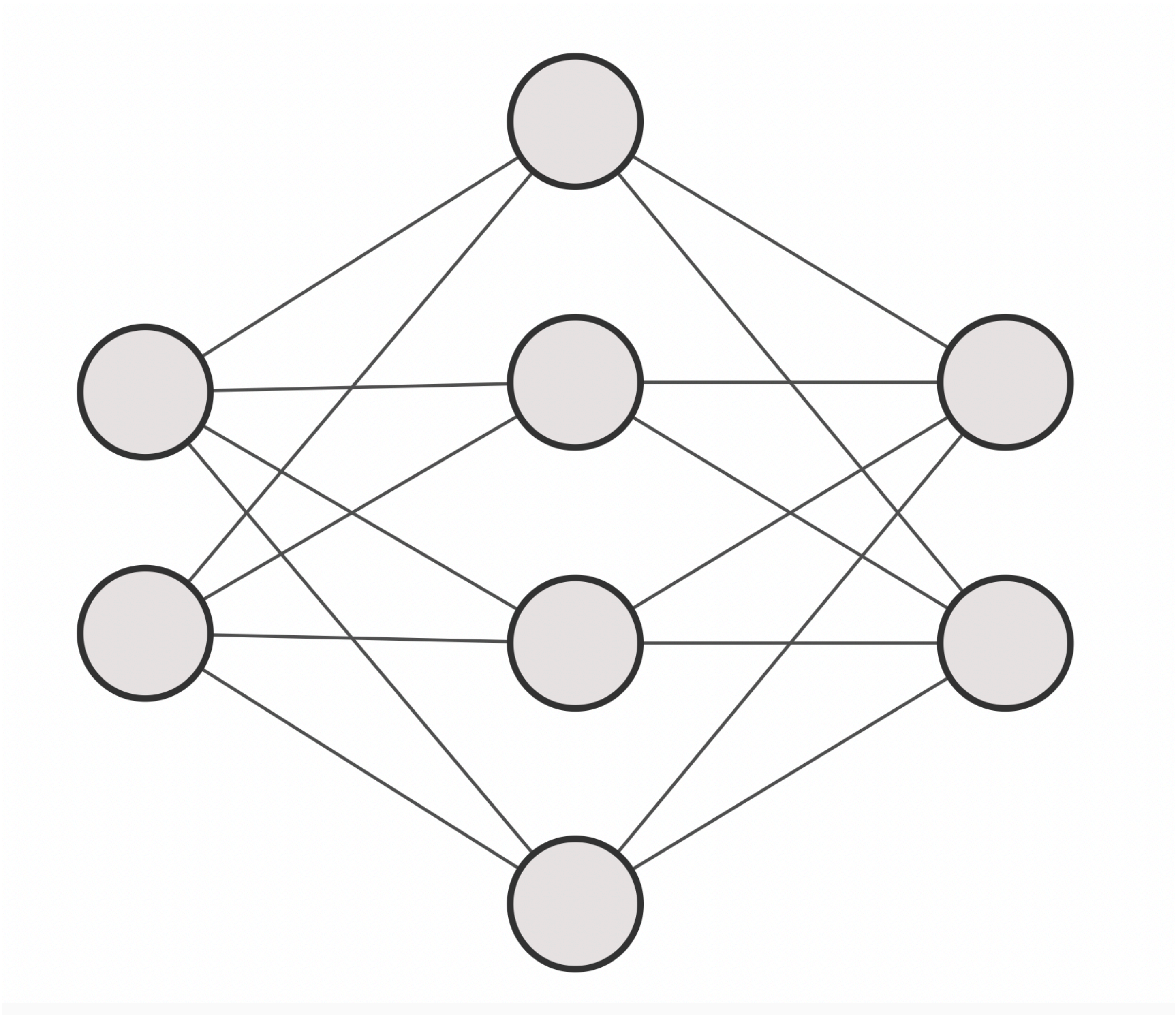


The Machine Learning

- Train neural network to predict $\nabla_x \log Q(x)$ using score matching techniques.



→



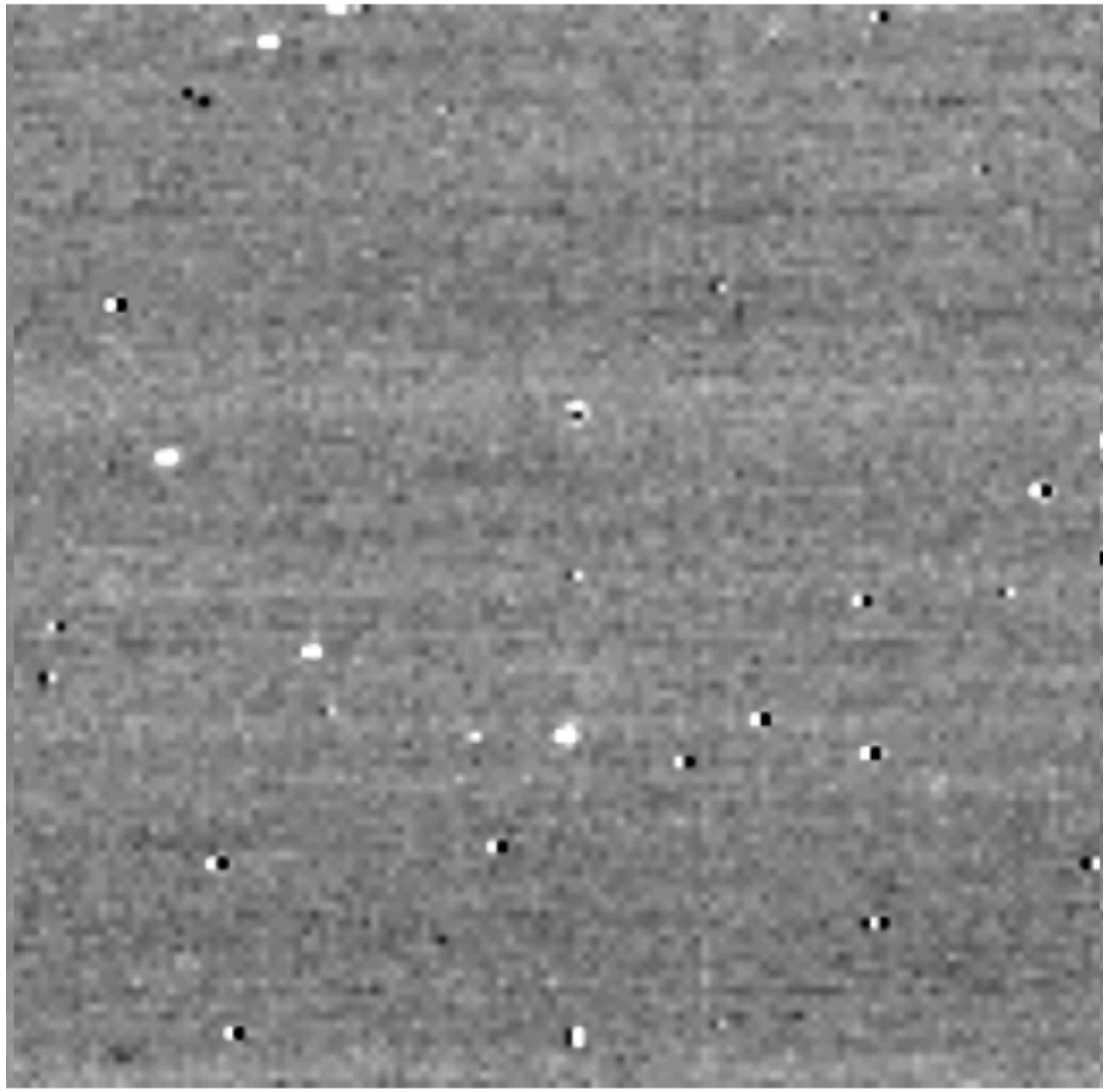
→ $\nabla_x \log Q(x)$

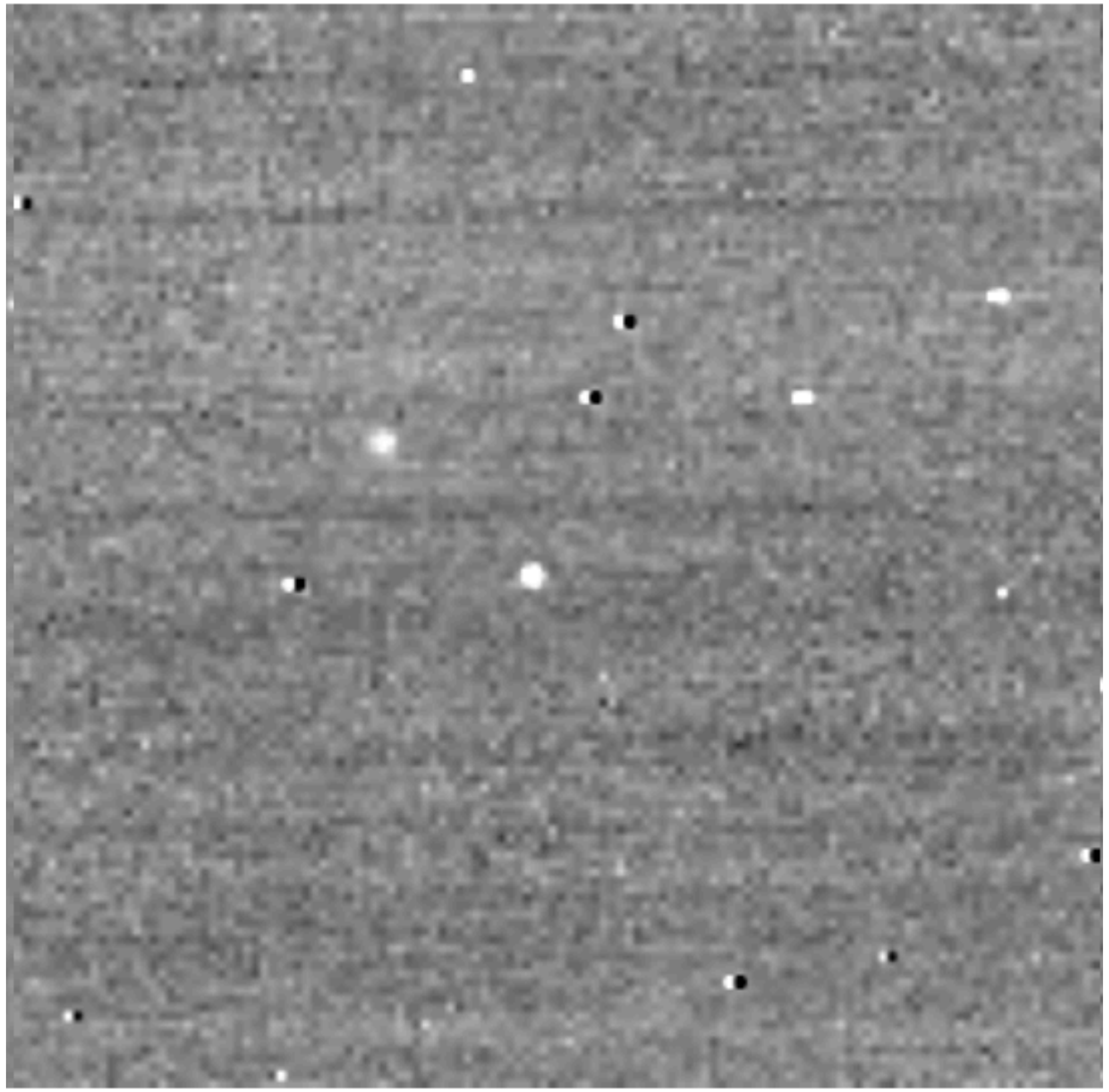
Machine Learning

Results

Noise Generation

- Can sample new noise. Which one is real?

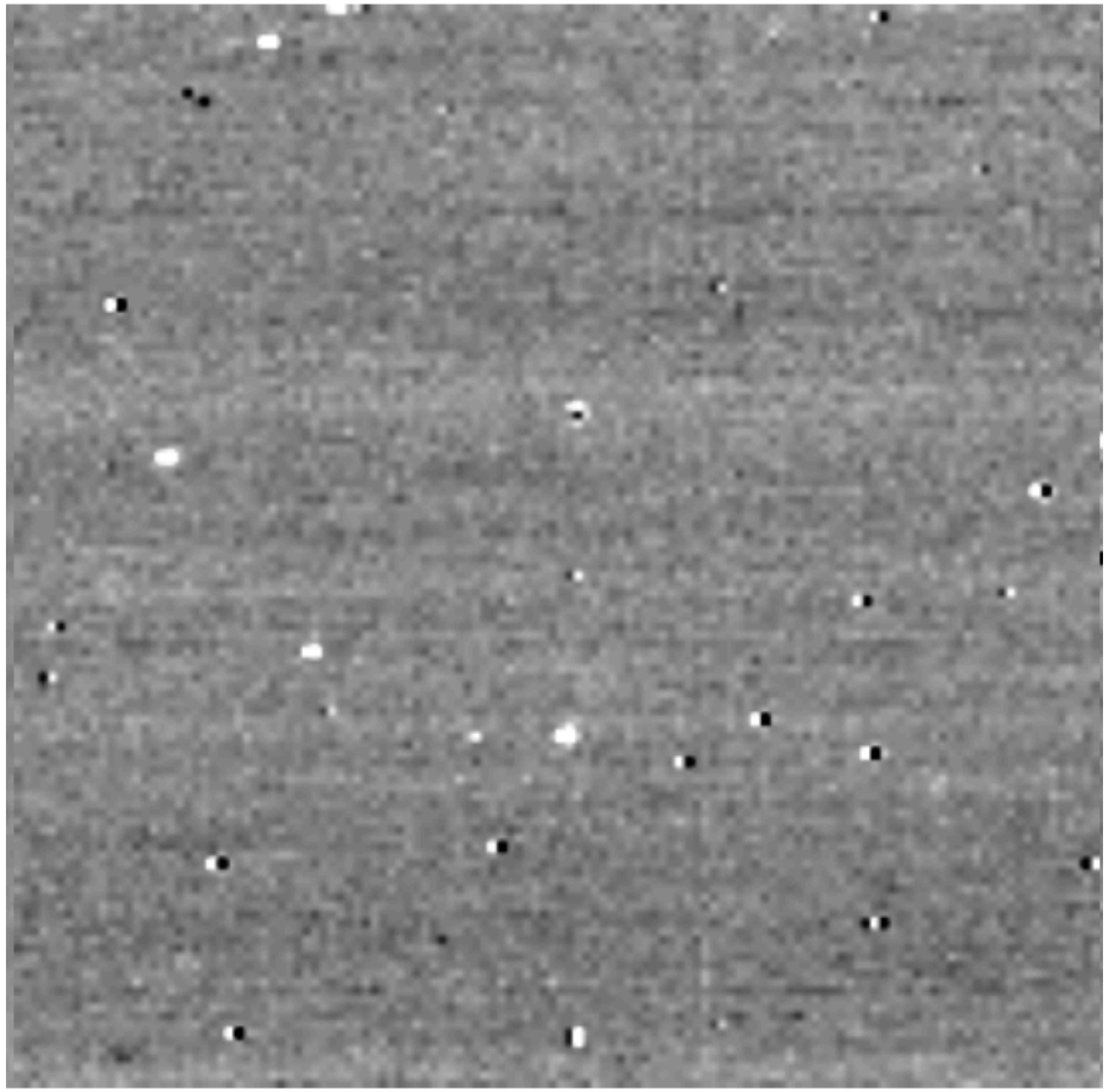




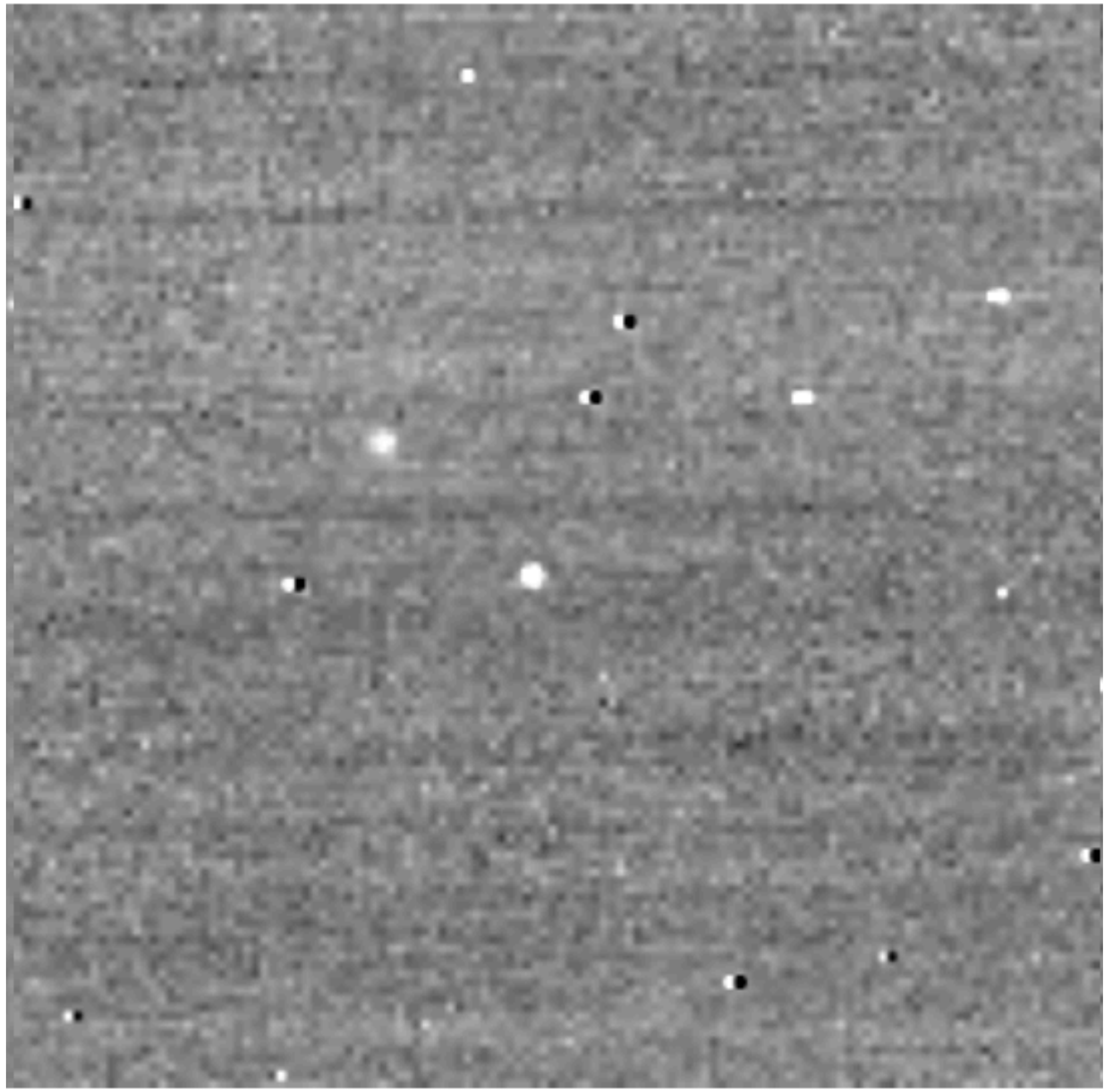
Noise Generation

- Can sample new noise. Which one is real?

Real

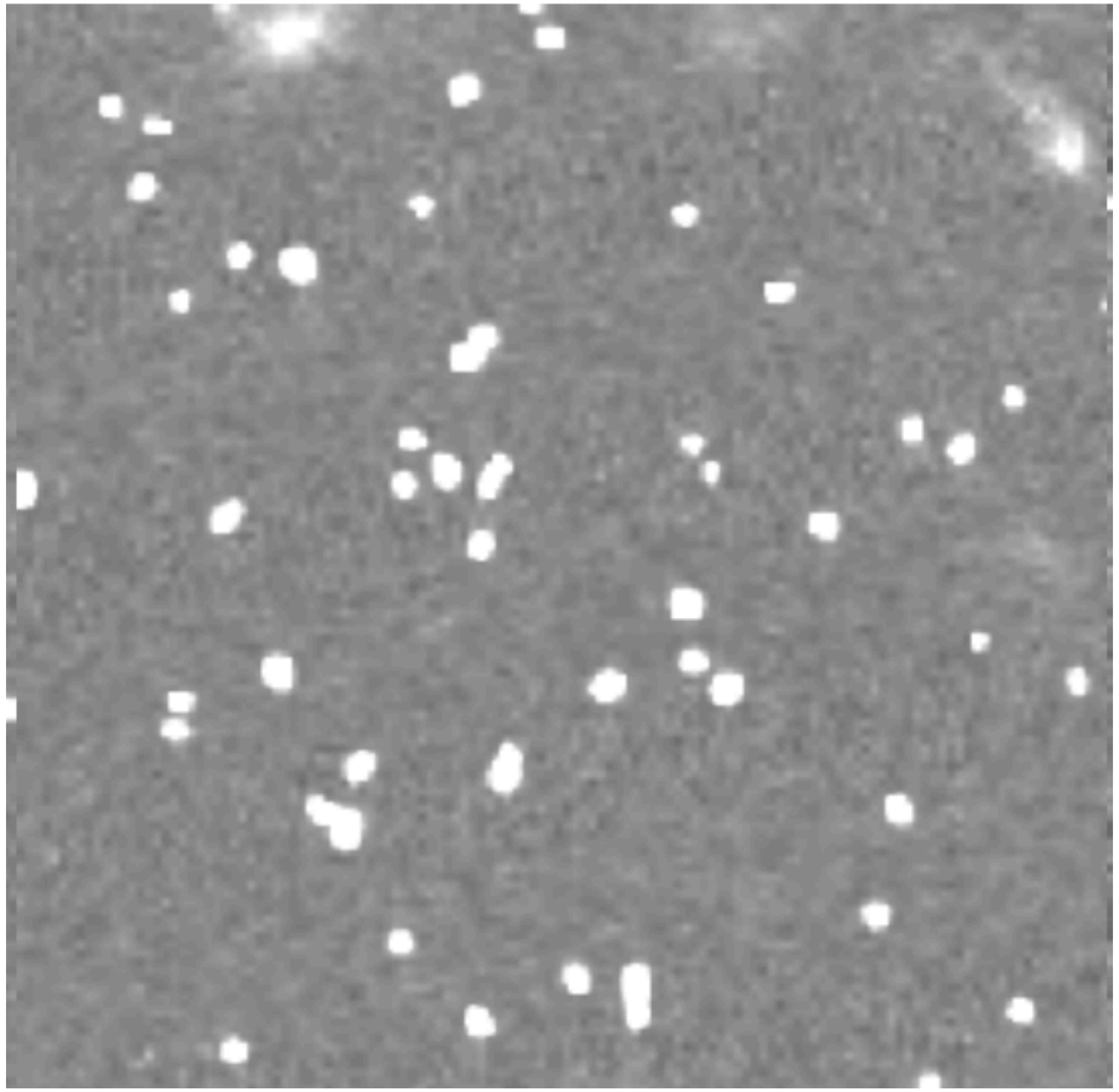


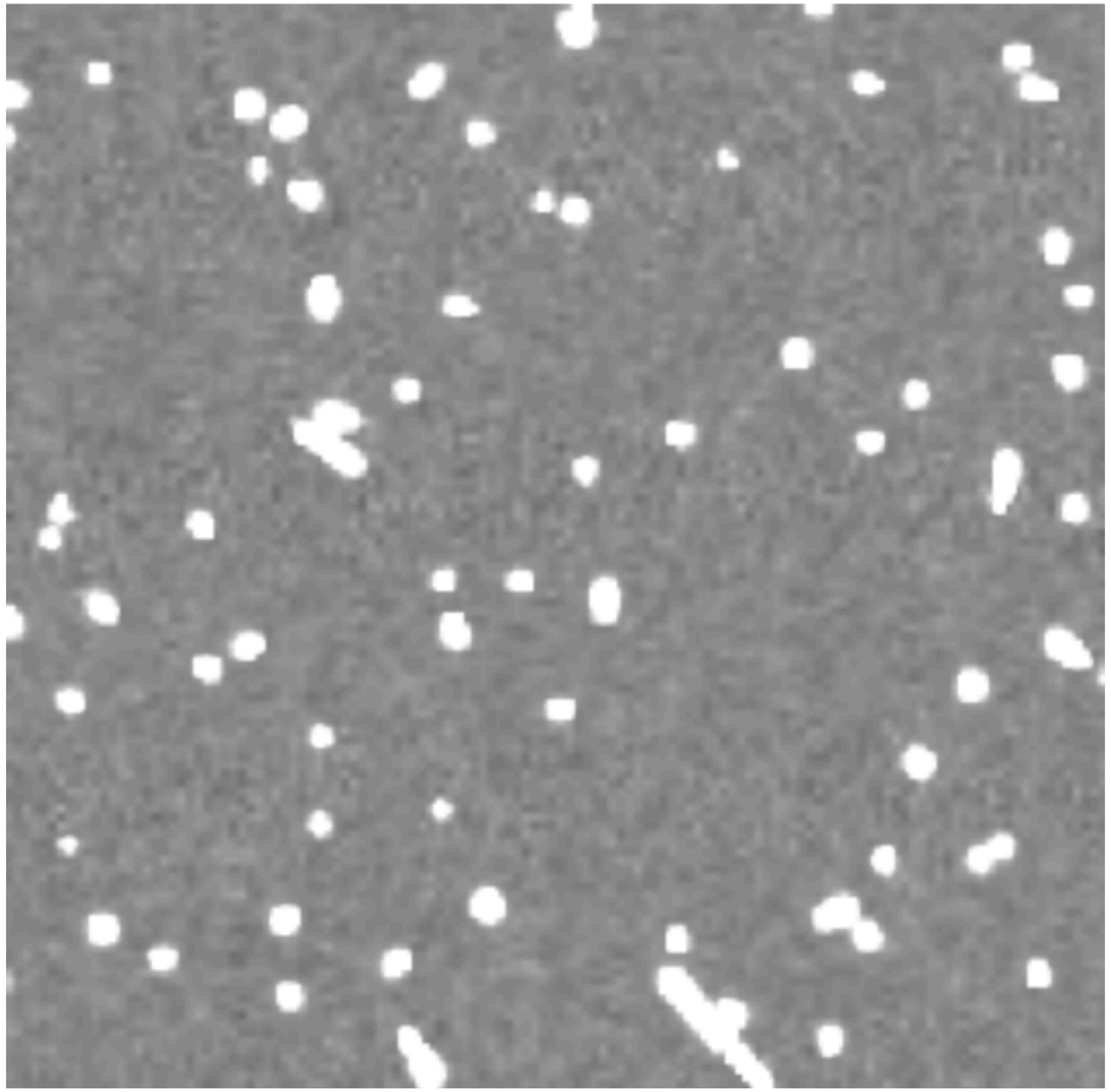
Fake (generated)



Noise Generation

- Can sample new noise. Which one is real?

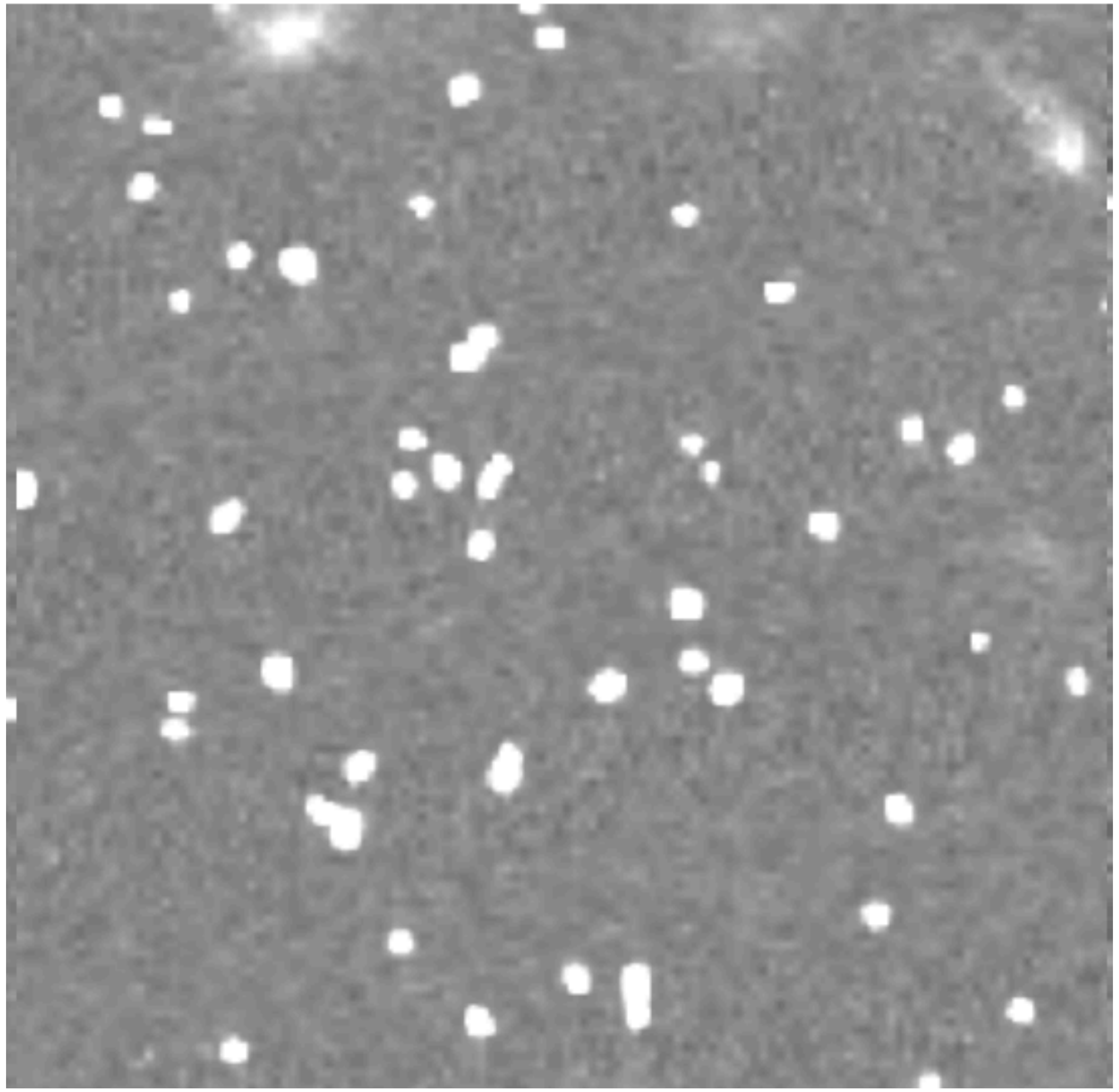




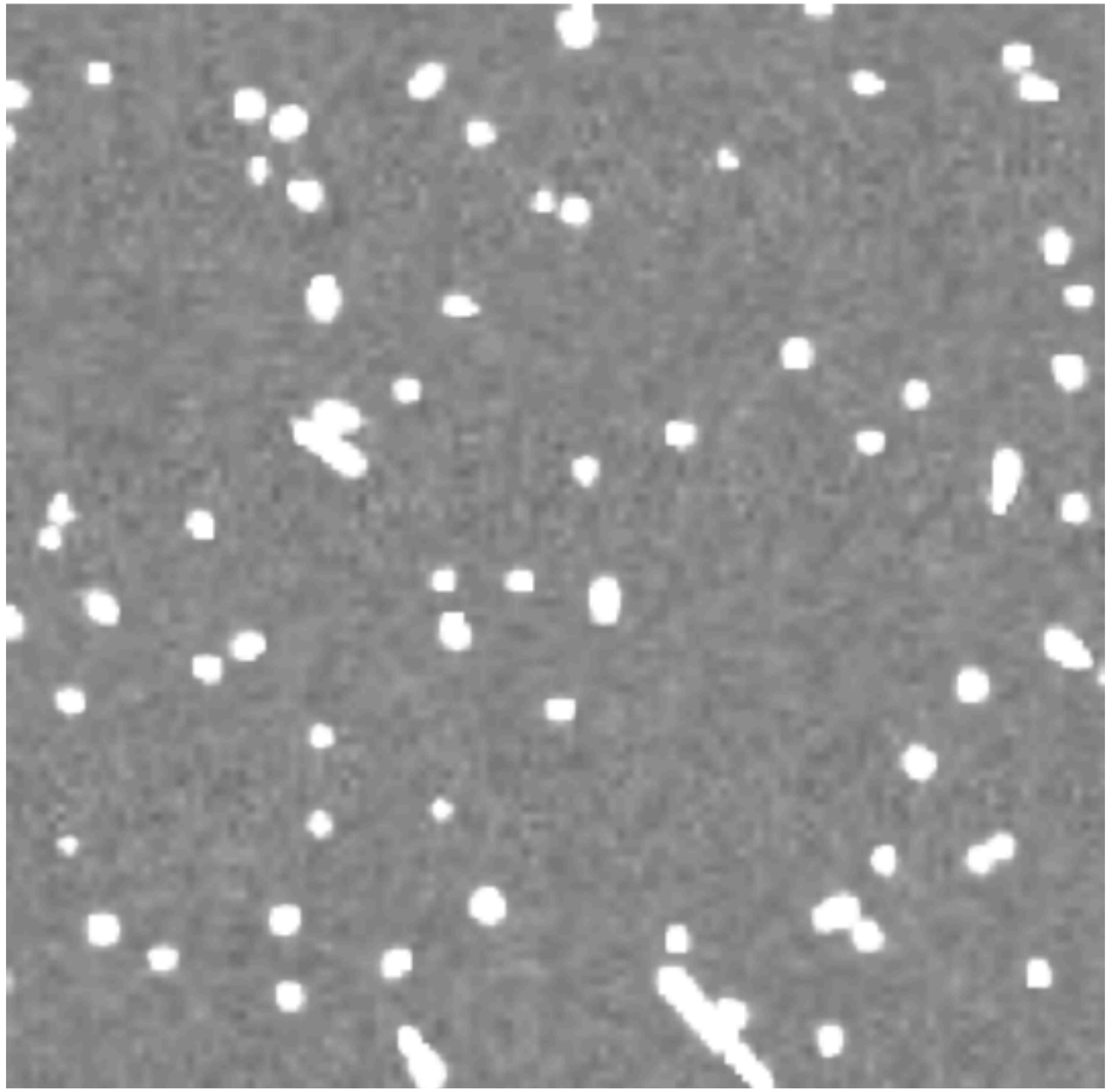
Noise Generation

- Can sample new noise. Which one is real?

Fake (generated)



Real



Besides sampling noise…

Inference!

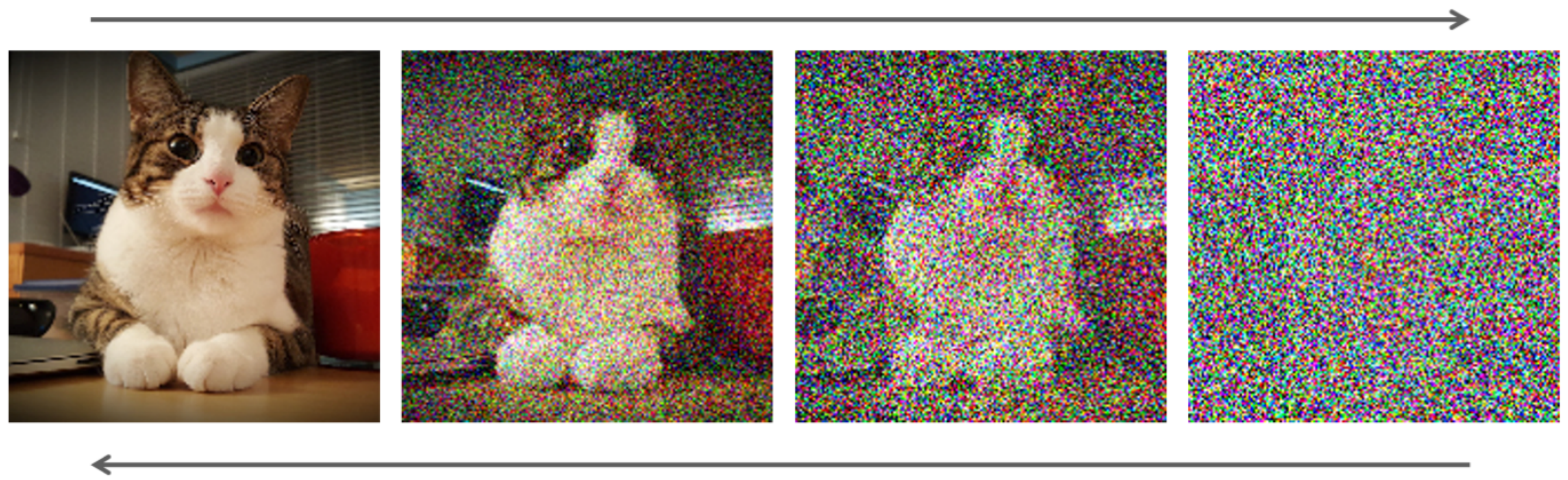
Posterior

$$p(\theta | d) = \frac{p(d | \theta)p(\theta)}{p(d)}$$

Likelihood

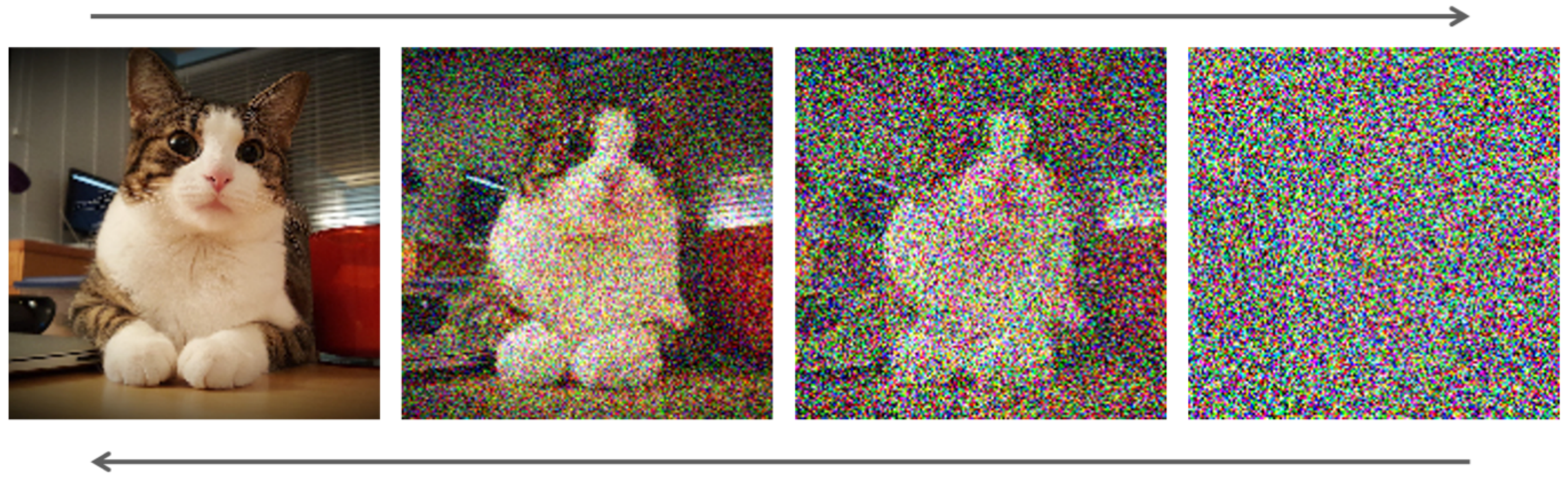
Posterior

$$p(\theta | d) = \frac{p(d | \theta) p(\theta)}{p(d)}$$



Posterior Likelihood

$$\nabla_\theta \log p(\theta | d) = \nabla_\theta \log p(d | \theta) + \nabla_\theta \log p(\theta)$$



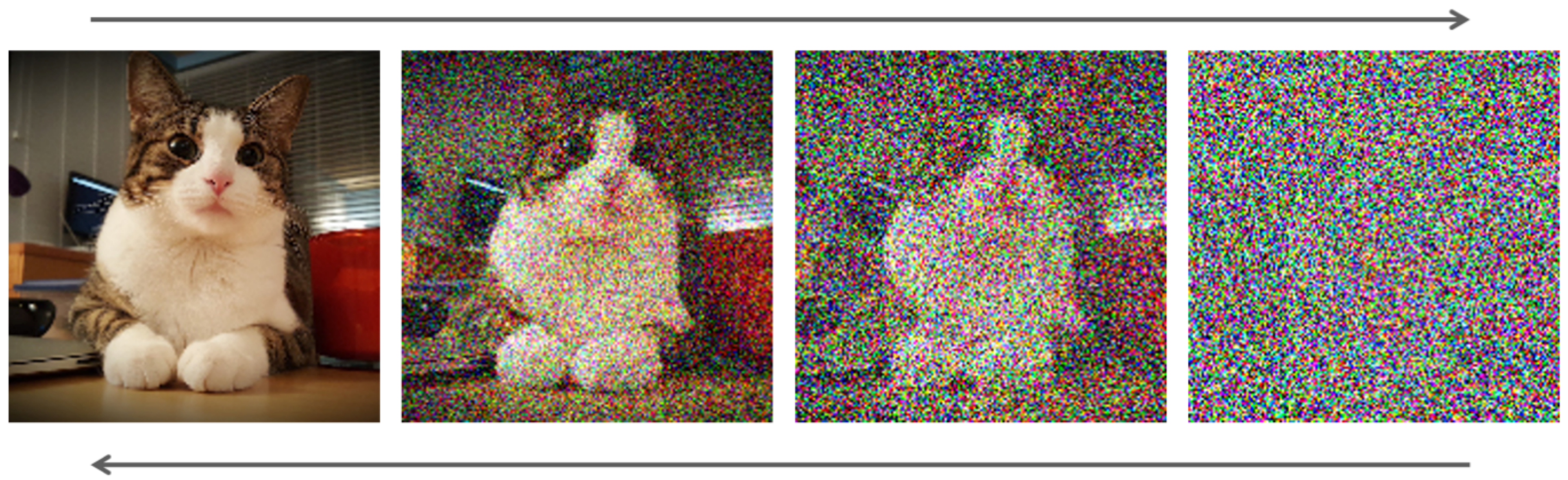
Posterior

Likelihood

$$\nabla_\theta \log p(\theta \mid d) = \nabla_\theta \log p(d \mid \theta) + \nabla_\theta \log p(\theta)$$

Learned Noise

$$\nabla_x \log Q(x)$$



Posterior

Likelihood

$$\nabla_\theta \log p(\theta \,|\, d) = \nabla_\theta \log p(d \,|\, \theta) + \nabla_\theta \log p(\theta)$$

SLIC Framework
(Score-based Likelihood Characterization)

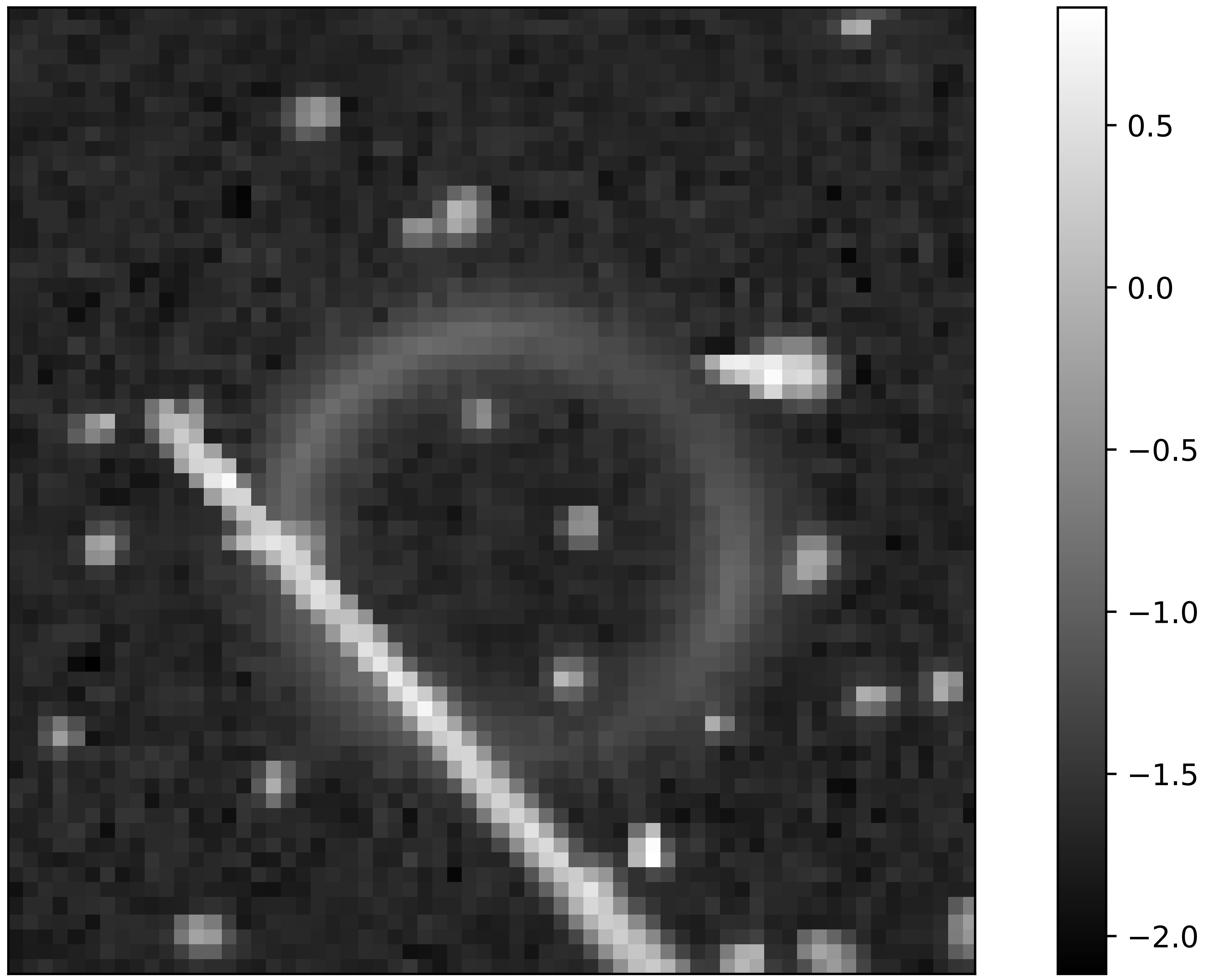
- Integrate learned noise distribution within well-defined Bayesian framework.
- Currently tested on additive noise $X = M(\theta) + N$
- $P(d|\theta) = Q(d - M(\theta)) \rightarrow \nabla_\theta \log Q(d - M(\theta)) = -\nabla \log Q \cdot \nabla_\theta M$

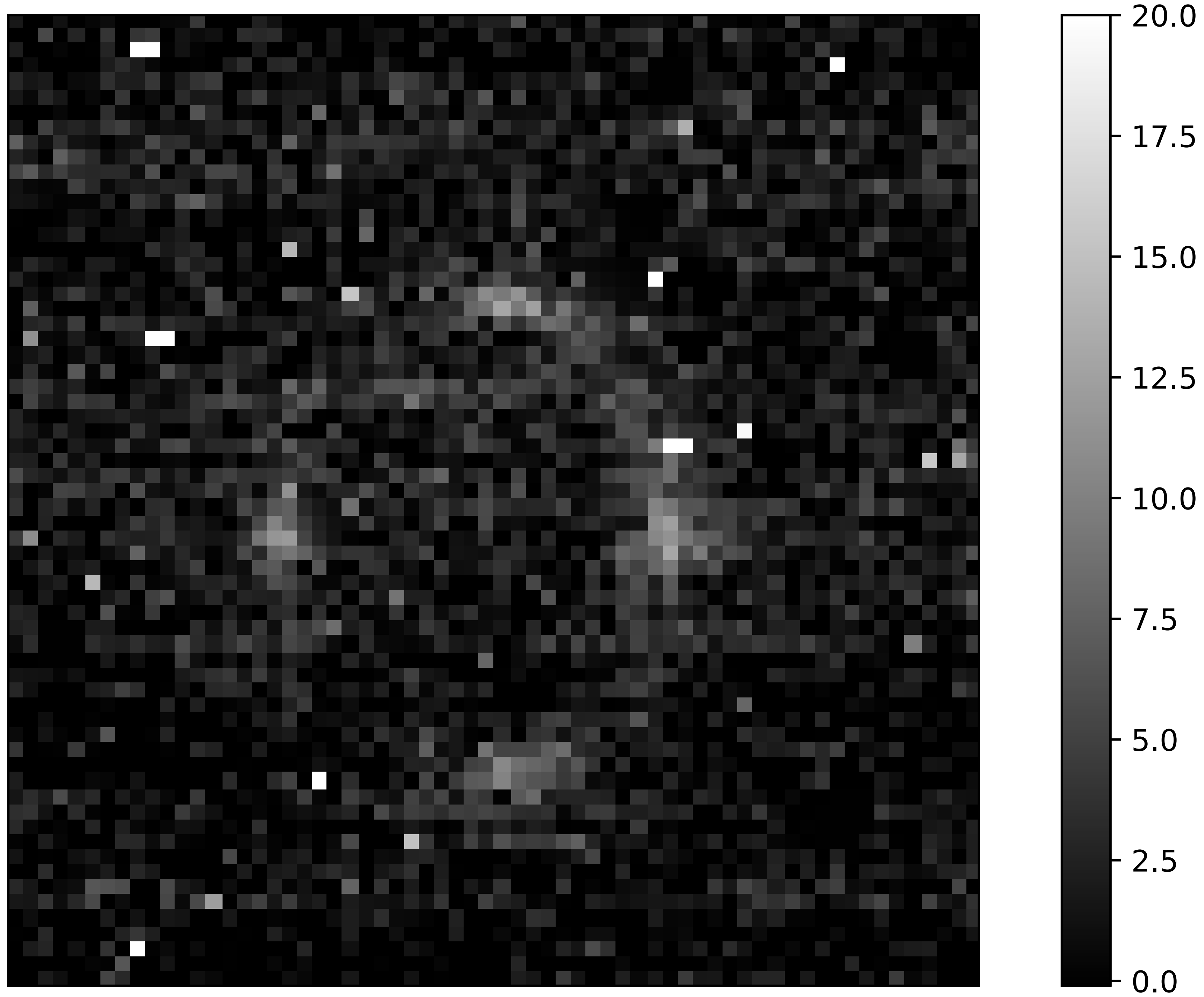
Our noise model!

SLIC Tests

SLIC Example

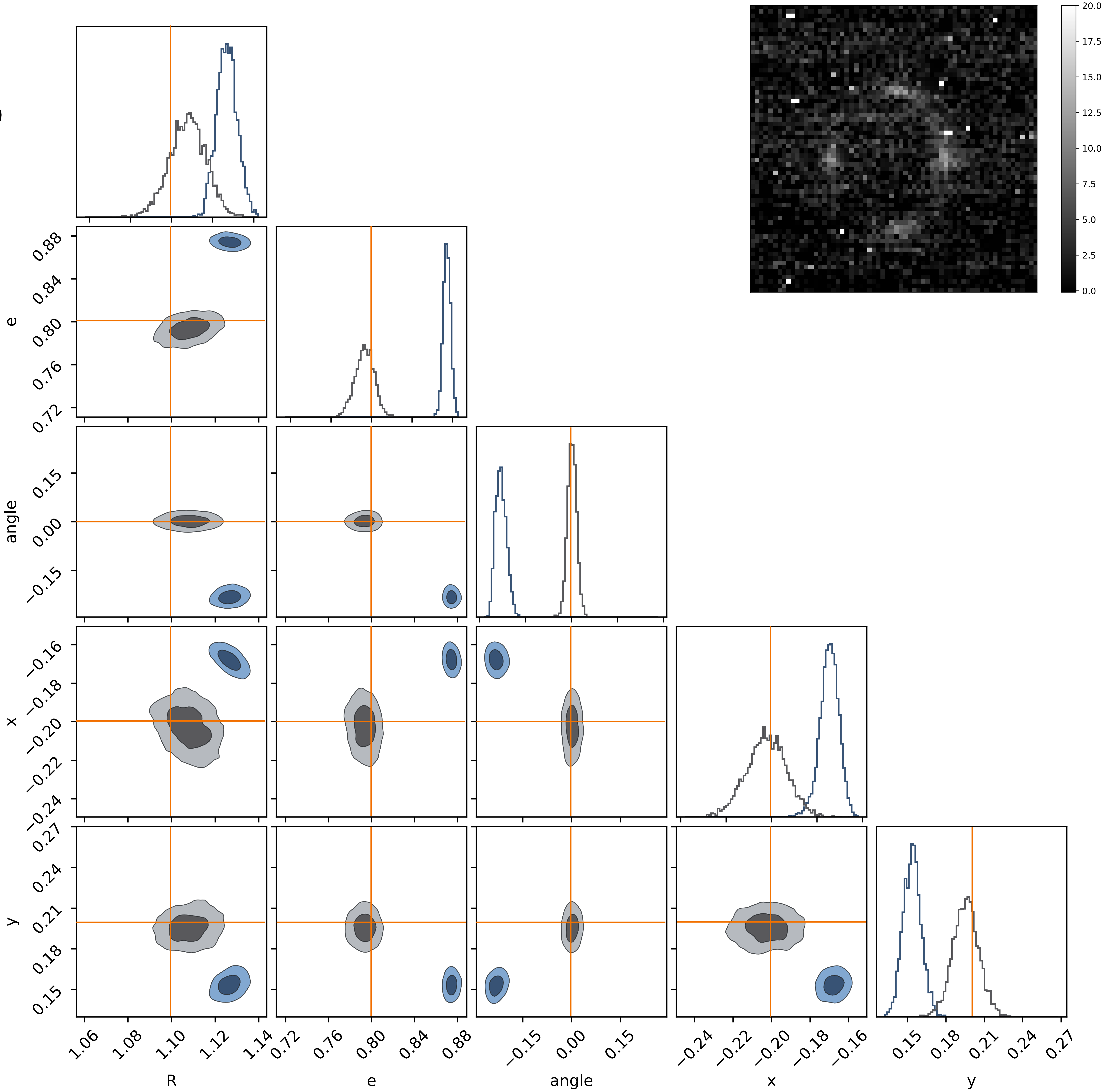
- Real noise + simulated strong lens test problem





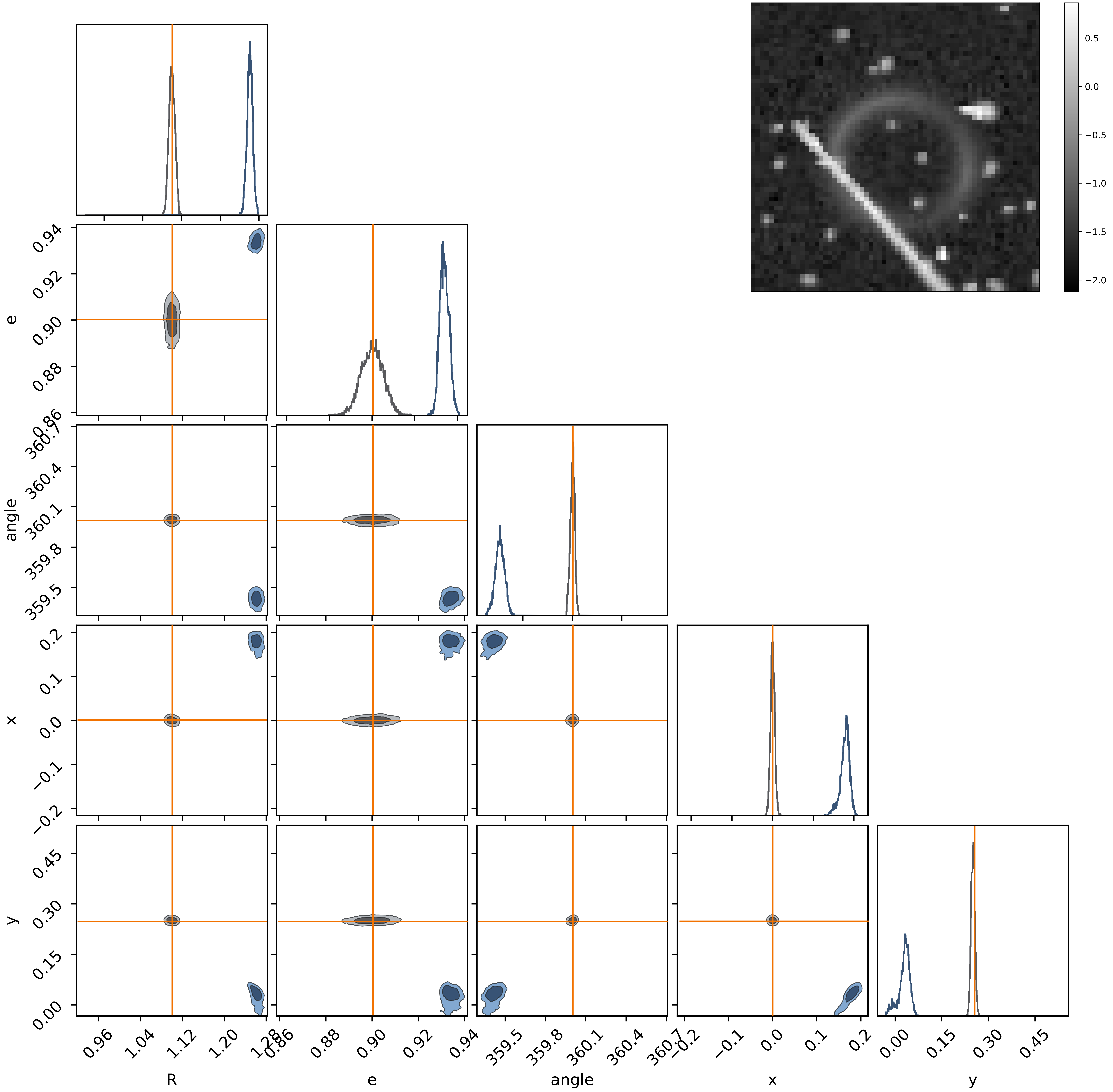
SLIC JWST Results

- We achieve accurate inference!
- Blue is Gaussian Likelihood
- Grey (middle blob) is SLIC.



SLIC HST Results

- We achieve accurate inference!
- Blue is Gaussian Likelihood
- Grey (middle blob) is SLIC.



Beyond Gaussian Noise: A Generalized Approach to Likelihood Analysis with non-Gaussian Noise

Ronan Legin,[1,2,3,*] Alexandre Adam,[1,2,3,*] Yashar Hezaveh,[1,2,3,4] and Laurence Perreault Levasseur[1,2,3,4]

[1] *Department of Physics, Université de Montréal, Montréal, Canada*
[2] *Ciela - Montreal Institute for Astrophysical Data Analysis and Machine Learning, Montréal, Canada*
[3] *Mila - Quebec Artificial Intelligence Institute, Montréal, Canada*
[4] *Center for Computational Astrophysics, Flatiron Institute, 162 5th Avenue, 10010, New York, NY, USA*



Conclusion

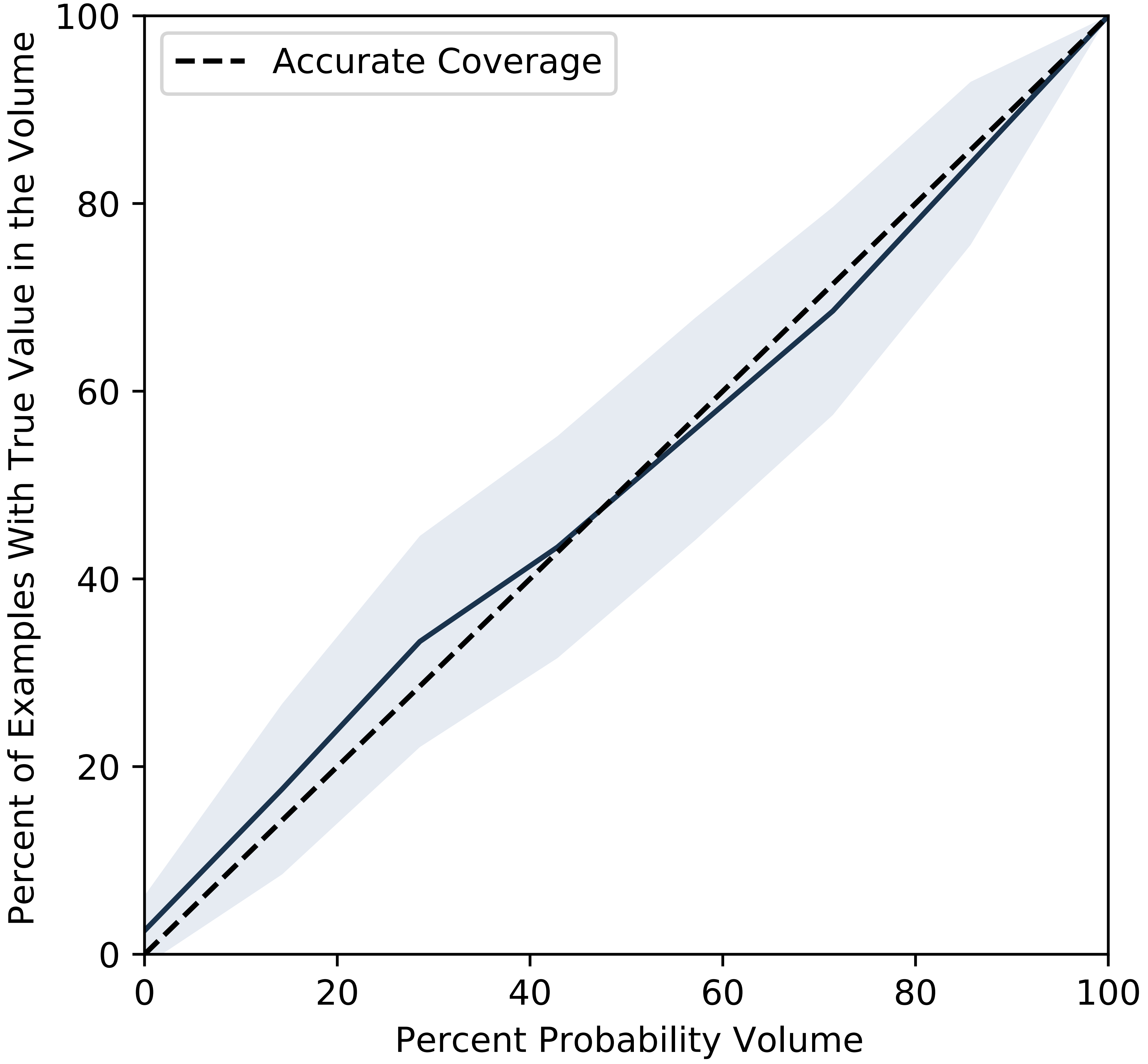
- Astrophysical noise is often non-Gaussian.
- We learn this noise to perform unbiased statistical inference.
- SLIC https://arxiv.org/abs/2302.03046

Thank you!

Extra slides

Coverage Tests

SLIC



The Machine Learning

- Score matching with transfer kernel $p(x_t \mid x_0) = \mathcal{N}(x_0, \sigma^2(t))$

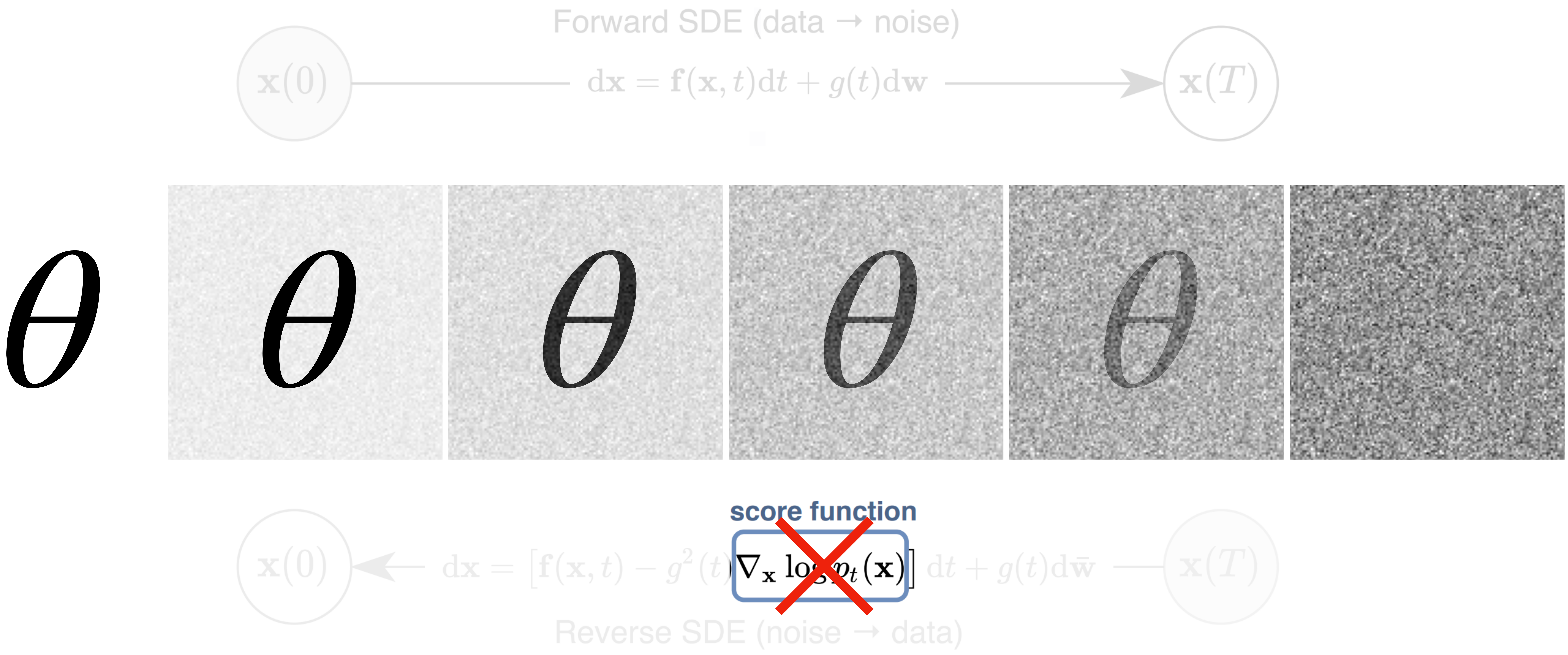
$$\| f_\eta(x_t, t) - \nabla_{x_t} \log p(x_t \mid x_0) \|^2$$

(Hyvärinen 2005; Vincent 2011; Song et al. 2020)

Posterior sampling

Likelihood — Prior

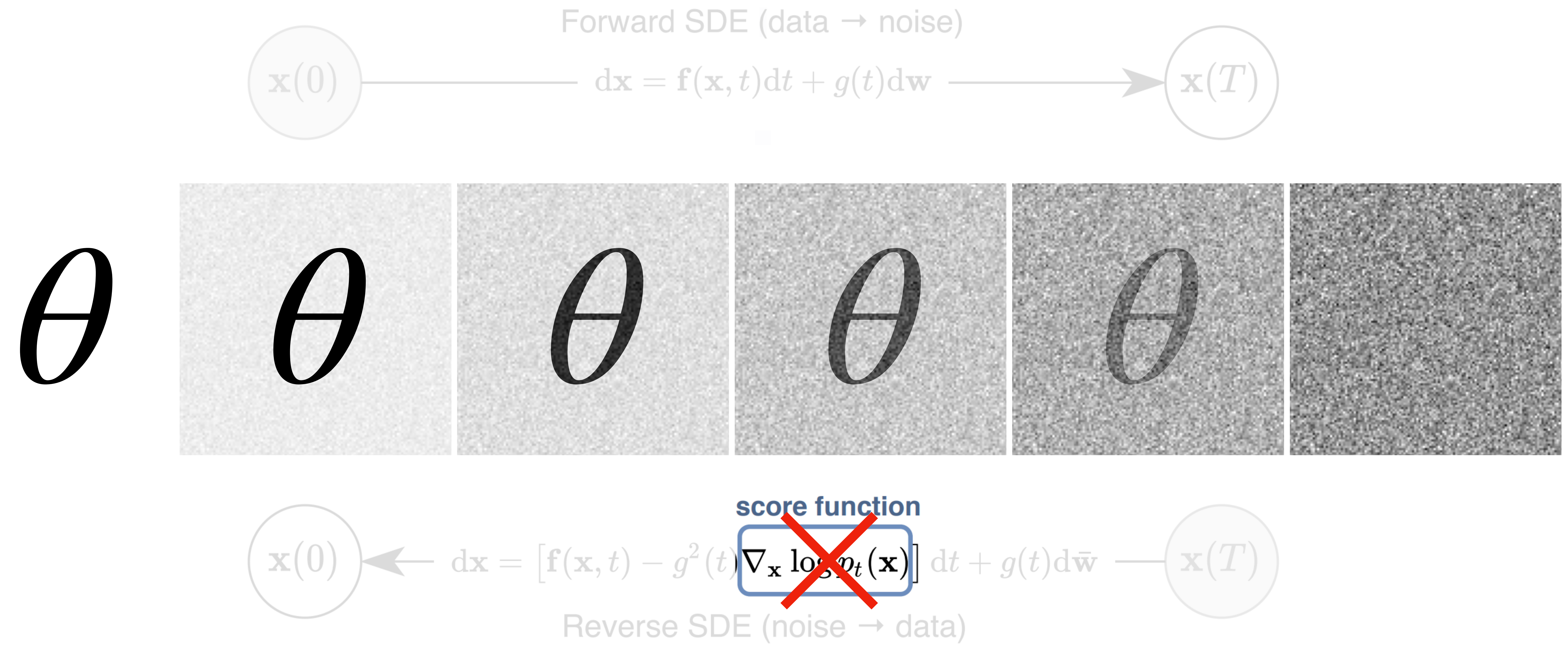
$$\nabla_\theta \log P_t(X \mid \theta) + \nabla_\theta \log P_t(\theta)$$



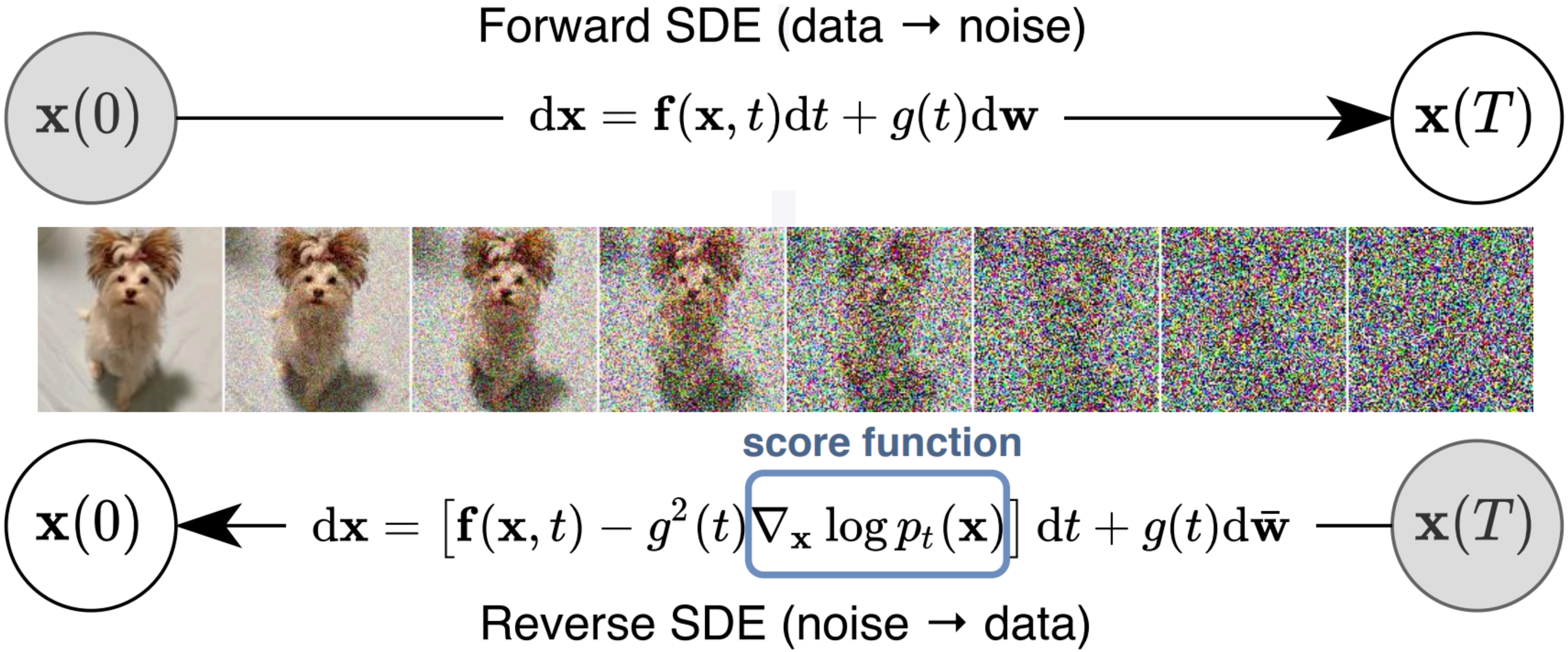
SLIC inference

Likelihood Prior

$$- \nabla_{X-M(\theta)} \log Q \cdot \nabla_\theta M(\theta) + \nabla_\theta \log P_t(\theta)$$



Score-based generative models



SLIC Framework
(Score-based Likelihood Characterization)

- Assume additive noise $X = M(\theta) + N$
- Given previous point, we can write likelihood $P(X \mid \theta) = Q(X - M(\theta))$
- $Q(X - M(\theta))$ is probability density of noise

SLIC trick

Decomposition with chain rule

- $\nabla_\theta \log Q(X - M(\theta)) = -\nabla_{X-M(\theta)} \log Q(X - M(\theta)) \cdot \nabla_\theta M(\theta)$
- This separates noise distribution Q from forward simulator M

SLIC trick

Decomposition with chain rule

- $\nabla_\theta \log Q(X - M(\theta)) = -\nabla_{X-M(\theta)} \log Q(X - M(\theta)) \cdot \nabla_\theta M(\theta)$
- This separates noise distribution Q from forward simulator M

Model $\nabla_{X-M(\theta)} \log Q(X - M(\theta))$ using score network!