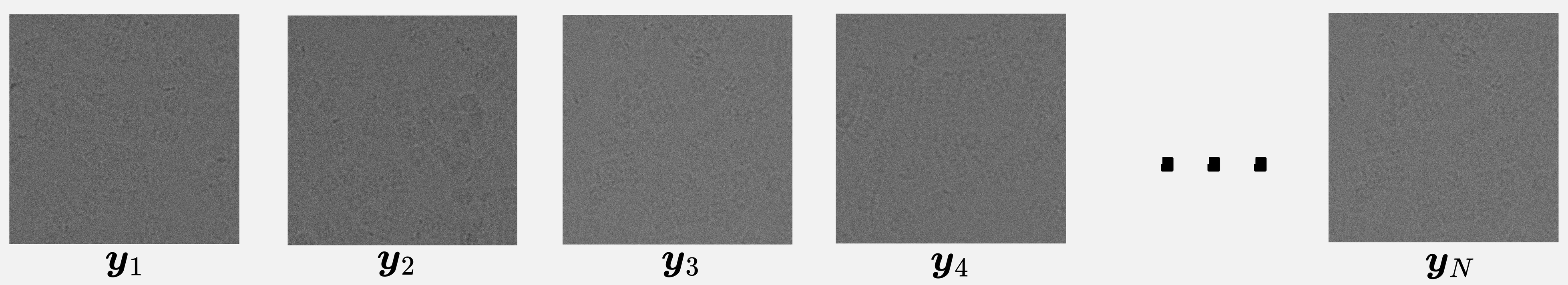
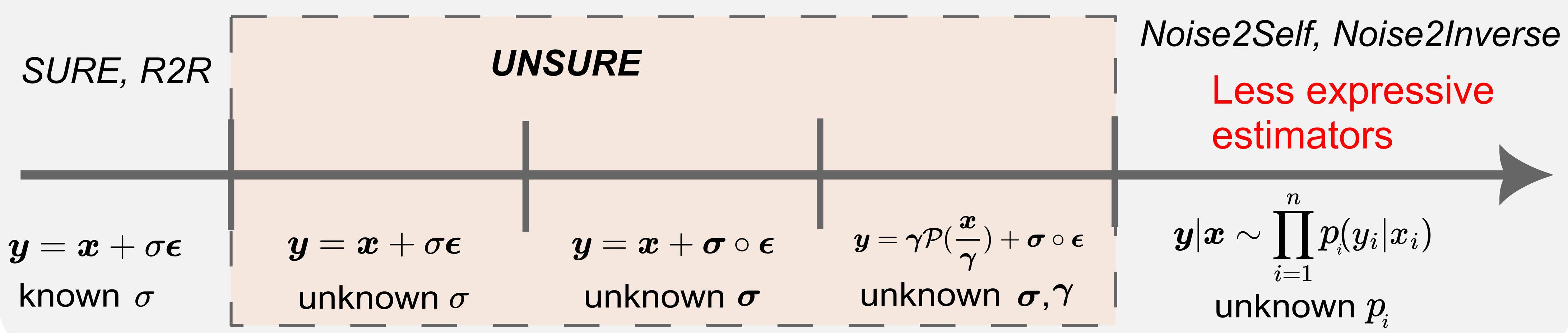
UNSURE: self-supervised learning with Unknown Noise level Stein's Unbiased Risk Estimate

Julián Tachella, Mike Davies and Laurent Jacques

Q. What can we learn from noisy images alone?



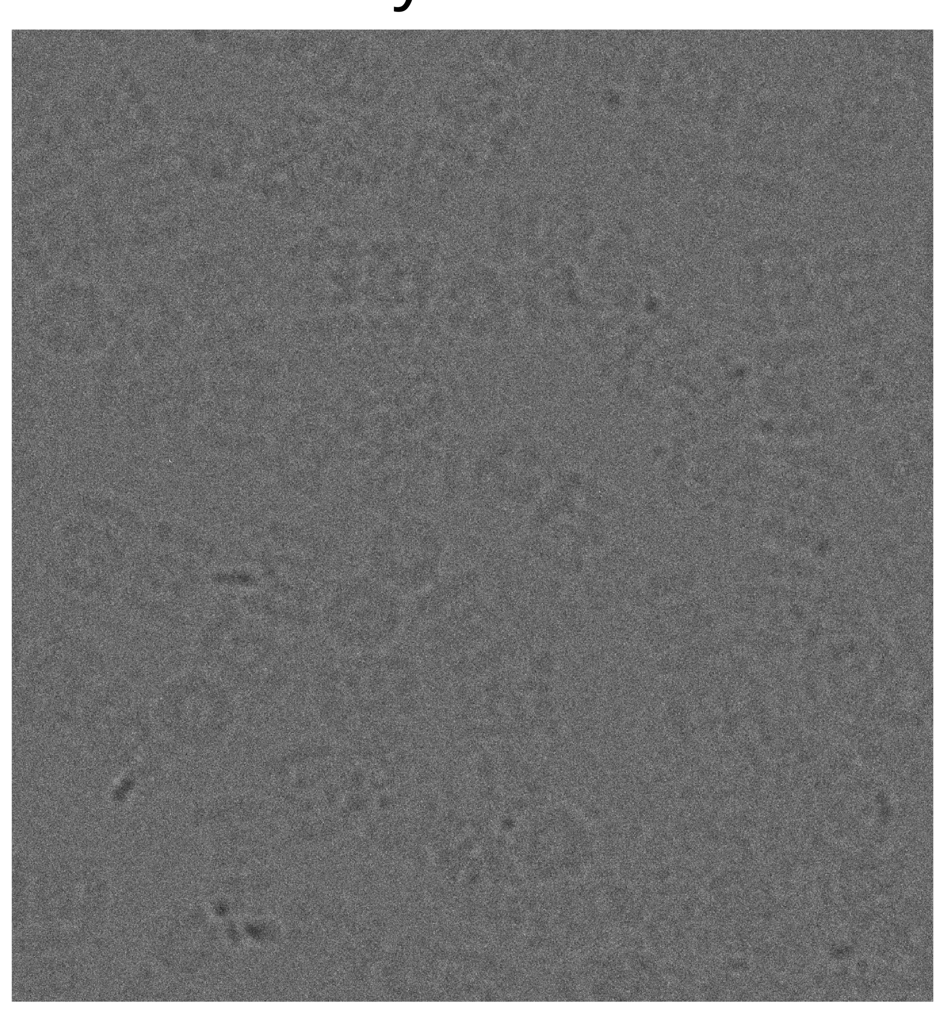
A. Depends on knowledge about the **noise distribution**



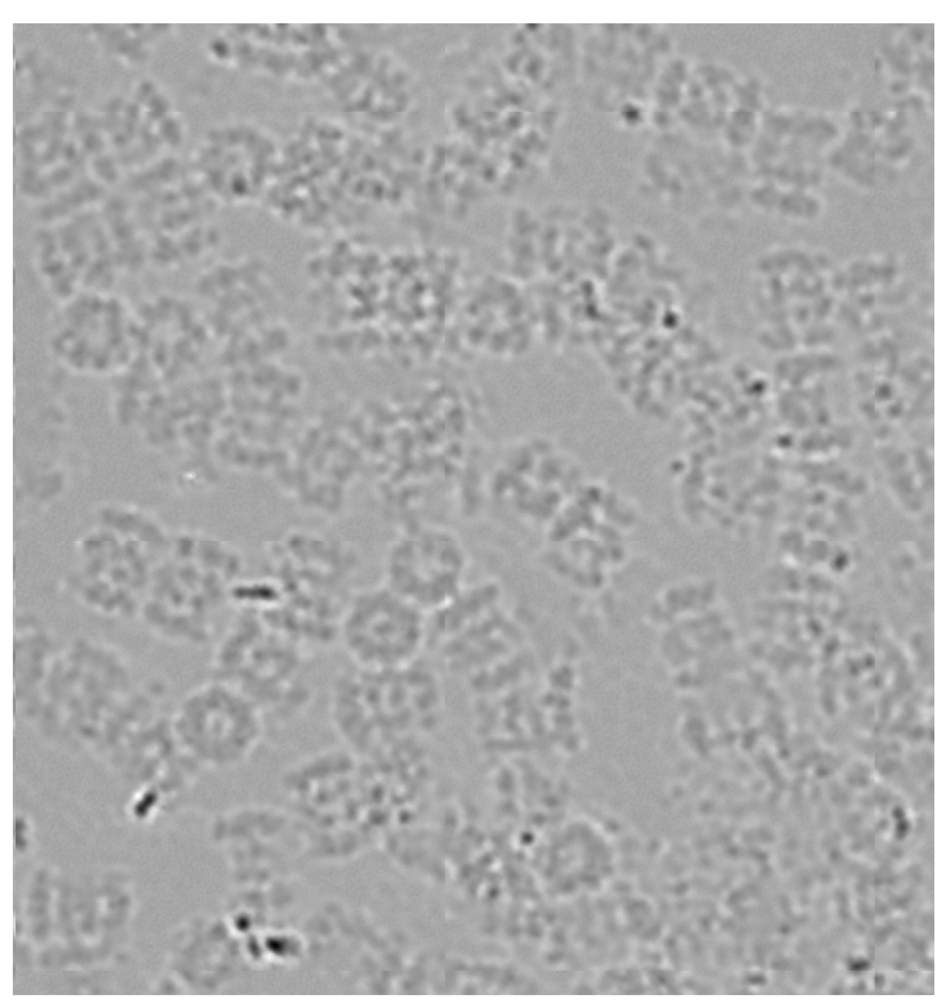
Experiments

PSNR [dB] in yellow

Real Cryo-EM data



UNSURE



Q. What loss should I minimize?

A. Consistency with constraints

$$\min_f \mathbb{E}_y \|f(y) - y\|^2 \text{ s.t. constraint } f = 0$$

a) Gaussian noise unknown σ

$$\sum_{i=1}^{n} \mathbb{E}_y \frac{\partial f_i}{\partial y_i}(y) = 0$$

b) Poisson Gaussian unknown (σ, γ)

$$\sum_{i=1}^{n} \mathbb{E}_y \frac{\partial f_i}{\partial y_i}(y) = 0 \qquad \sum_{i=1}^{n} y_i \frac{\partial f_i}{\partial y_i}(y) = 0$$

c) Gaussian noise unknown Σ

Possible covariances: $\{\Sigma = \sum_{j=1}^{J} \eta_j \Psi_j,\ \eta \in \mathbb{R}^J\}$

$$\mathbb{E}_y \operatorname{tr}(\Psi_j \nabla f(y)) = 0 \text{ for } j = 1, \dots, J$$

d) Other models ... (see paper)

Q. Why these losses?

A. Because of SURE

$$\min_f \mathbb{E}_y \|f(y) - y\|^2 + 2\sigma^2 \sum_{i=1}^{n} \cancel{\mathbb{E}_y \frac{\partial f}{\partial y_i}(y)}$$

Q. What's the performance?

A. Almost as supervised

a) Generalization of Tweedie

$$f(y) = y + \frac{n\nabla \log p(y)}{\mathbb{E}_y \|\nabla \log p(y)\|^2}$$

b) Expected performance

$$\mathbb{E}_{x,y} \frac{1}{n}\|f(y) - x\|^2 = \text{MMSE} + \sigma^2 \sum_{j=2}^{\infty} \left(\frac{\text{MMSE}}{\sigma^2}\right)^j.$$

c) Upper bound noise level

$$\frac{n}{\mathbb{E}_y \|\nabla \log p(y)\|^2} = \frac{\sigma^2}{1 - \frac{\text{MMSE}}{\sigma^2}} \geq \sigma^2$$

Q. Implementation?

A. Lagrange multipliers

$$\min_f \max_\eta \mathbb{E}_y \|f(y) - y\|^2 + 2\operatorname{tr}\left(\Sigma_\eta \frac{\partial f}{\partial y}(y)\right)$$

Divergence approximation

$b \sim \mathcal{N}(0, I) \qquad 0 < \tau \ll 1$

$$\operatorname{tr}\left(\Sigma_\eta \frac{\partial f}{\partial y}(y)\right) \approx \frac{(\Sigma_\eta b)^\top}{\tau}\left(f(y + \tau b) - f(y)\right)$$

Algorithm 1 UNSURE loss.

Require: step size α, momentum μ
residual $\leftarrow \|f_\theta(y) - y\|^2$
$b \sim \mathcal{N}(0, I)$
div $\leftarrow \frac{(\Sigma_\eta b)^\top}{\tau}(f_\theta(y + \tau b) - f_\theta(y))$
loss $\leftarrow$ residual + div
$g \leftarrow \mu g + (1 - \mu)\frac{\partial \text{div}}{\partial \eta}$
$\eta \leftarrow \eta + \alpha g$
return loss(θ)

Added lines vs SURE