École de technologie supérieure
Département de génie logiciel et des technologies de l'information
Neuro-IX

Presentation of a Paper For Lab-Meet

DDM^2 : SELF-SUPERVISED DIFFUSION MRI DENOISING WITH GENERATIVE DIFFUSION MODELS

Tiange Xiang, Mahmut Yurt, Ali B Syed, Kawin Setsompop & Akshay Chaudhari
Stanford University

Published as a conference paper at ICLR 2023

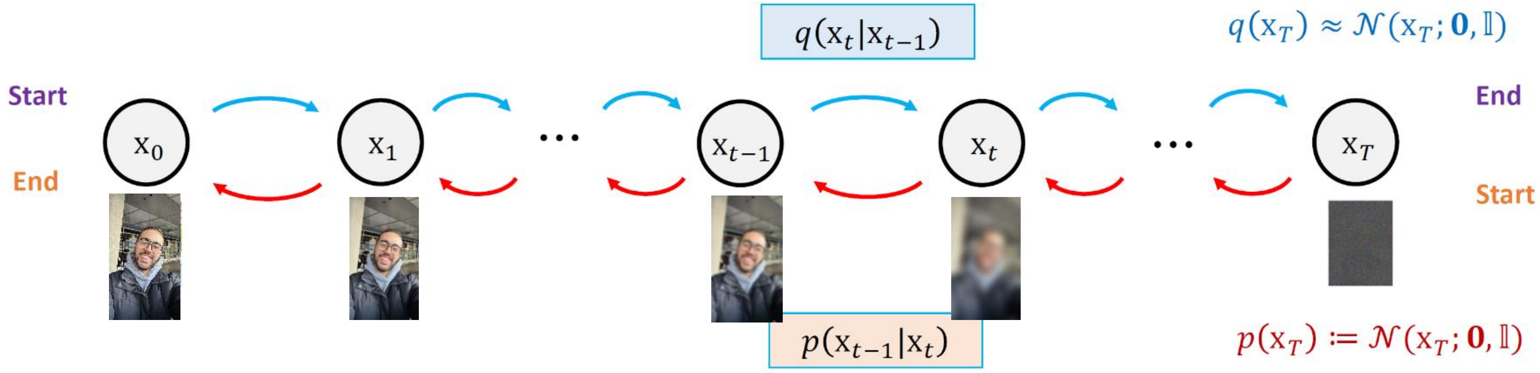
Prepared by:
Ahmed REKIK

Supervised by:
M.Sylvain BOUIX

Session Automne 2025

Prerequisites / Recap : Diffusion Models

- Generative models that can create new data.
- Process: gradually add noise until the signal is destroyed.
- Learning: train a network to reverse the process and reconstruct data.



1/ Forward Process (Diffusion / Noise Addition)

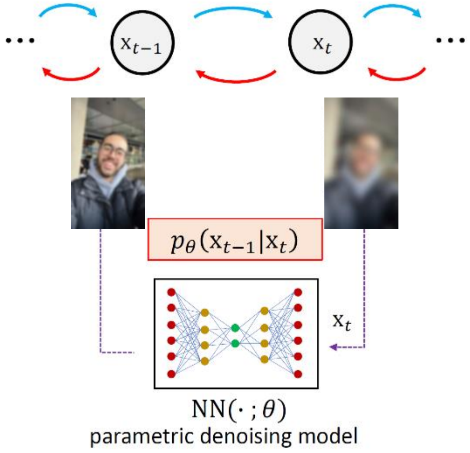
- Start from a clean image x_0.
- Gradually add Gaussian noise across T steps : $q(x_t \mid x_{t-1}) = \mathcal{N}(x_t; \sqrt{1-\beta_t}x_{t-1}, \beta_t I)$
- After many steps → the image becomes pure noise.

2/ Reverse Process (Denoising / Generation)

- Train a neural network F_θ to predict and remove the noise at each step.
- Sampling: start from noise and progressively denoise → generate a realistic image.

3/ Training Objective

- Learn to predict the added noise ϵ: $L(\theta) = \mathbb{E}_{x_0,t,\epsilon}\left[\|\epsilon - \epsilon_\theta(x_t, t)\|^2\right]$

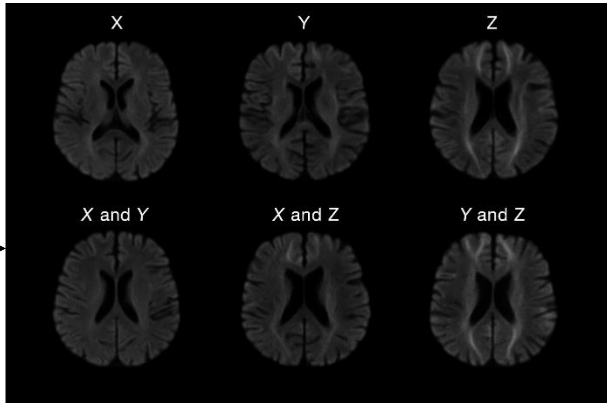


Context:

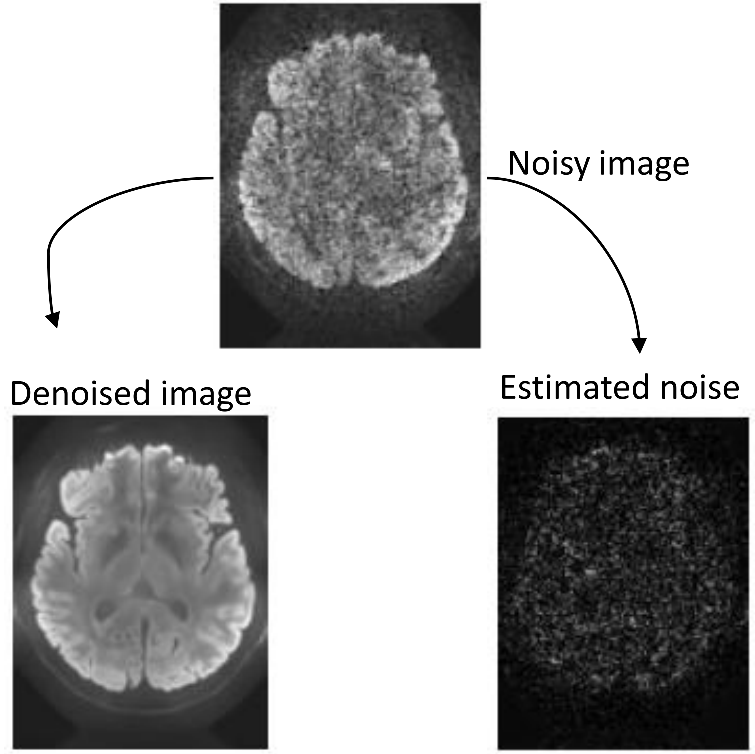
Diffusion MRI and the issue of noise

- **MRI:** A non-invasive medical imaging modality, essential for diagnosis.

- **Diffusion MRI (dMRI):**
  - Allows the analysis of tissue microstructure (brain, oncology).
  - Produces 4D images: 3D spatial + diffusion directions.

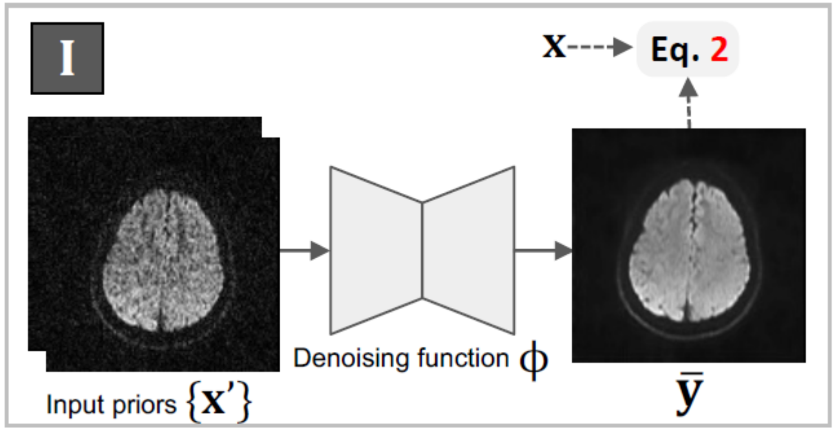


- **Problem ;**
  - **Low SNR** ($= \frac{\text{Useful signal amplitude}}{\text{Noise standard deviation}}$) → diagnostic losses.
  - Long acquisition times → patient discomfort + higher costs.
  - Difficult to acquire **paired low/high SNR** scans in clinical practice.
  - **Target Solution** :
    - **Self-supervised** denoising (no ground truth required).
    - Combination of **noise statistical** models and **diffusion models**.
    - Robust **generalization** across diverse MRI protocols.



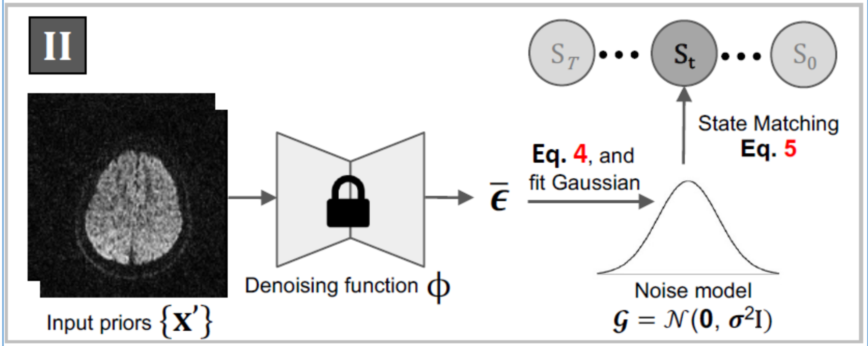
Vue d'ensemble de la méthode : DDM²

1. **Stage I** – Noise Model Learning

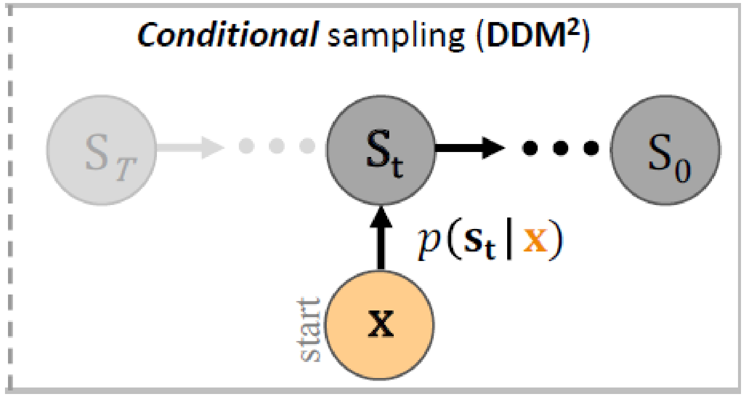


Learning a function Φ to estimate noise.

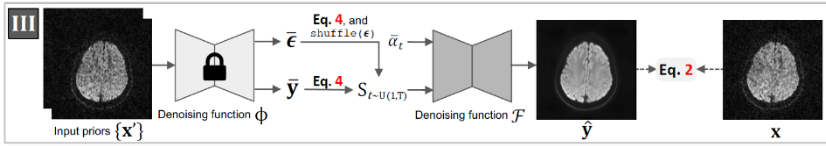
2. **Stage II** – Markov State Matching



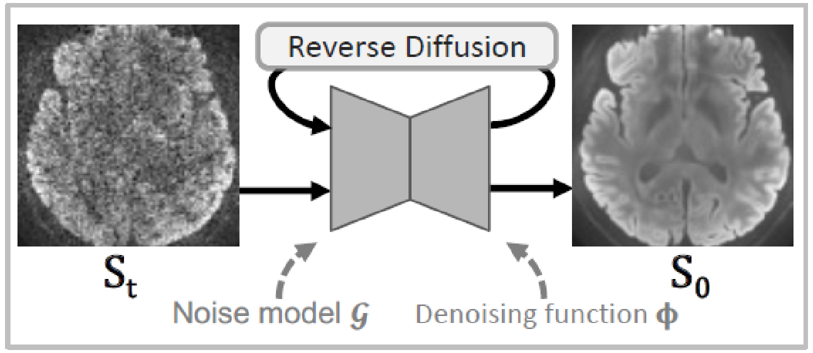
Associating each noisy image with an intermediate state of the diffusion Markov chain.



3. **Stage III** – Diffusion Model Training



Generation of clean images using a diffusion model F.



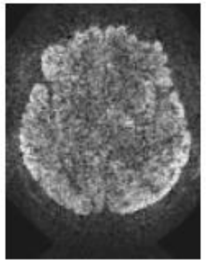
DDM^2 : Stage I

Noise Model Learning

Physical acquisition :

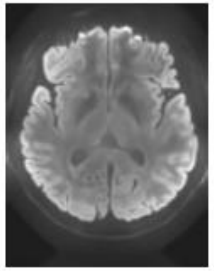
$$\mathbf{x} = \lambda_1 \mathbf{y} + \epsilon$$

Noisy image



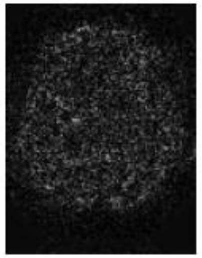
$= \lambda_1 \times$

Denoised image



$+$

Estimated noise $\varepsilon \sim \mathcal{N}(0, \lambda_2^2 I)$



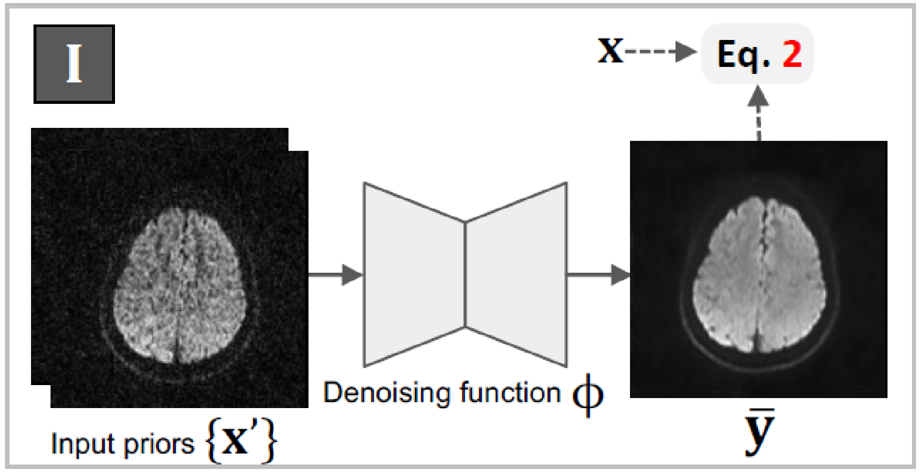
Predictive model $\mathbf{\Phi}$ (denoising function with: $\lambda_1 = 1$)

- ❑ Principle of J-Invariance :

A **denoising network** can be trained using only **noisy images**, by predicting a **clean approximation** $\bar{\mathbf{y}}$ from the other noisy slices {x′}.

$$\bar{\varepsilon} = x - \bar{y}$$

→ Residual noise estimated

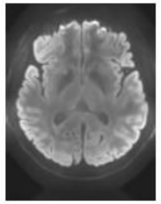


$$y \approx \bar{y} = \Phi(\{x'\})$$

$$\mathcal{L}(\Phi(\mathbf{x}'), \mathbf{x}) = ||\Phi(\mathbf{x}') - \mathbf{x}||^2 \approx ||\Phi(\mathbf{x}') - \mathbf{y}||^2 + \text{const.}$$

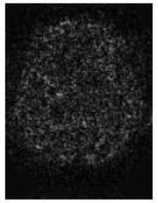
- ❑ Résultats :

Denoised image



$+$

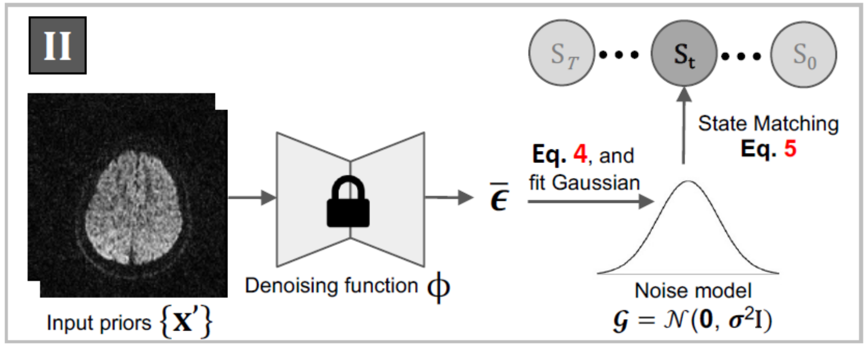
Estimated noise



Stage II

DDM^2 : Stage II
Markov State Matching

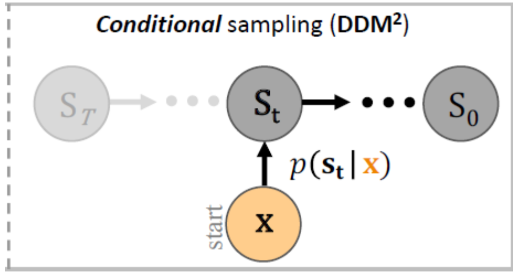
1. Centering of residual noise : $\bar{\varepsilon} := \bar{\varepsilon} - \mu_{\bar{\varepsilon}}$, $\bar{y} := \bar{y} + \frac{\lambda_2 \mu_{\bar{\varepsilon}}}{\lambda_1}$ → $\bar{\varepsilon} \sim \mathcal{N}(0, \sigma^2 I)$



2. Matching to a state in the diffusion chain :

- Compare the measured noise $\boldsymbol{\sigma}$ with the posteriors $\boldsymbol{p(S_t)}$, associated with the noise level $\boldsymbol{\beta_t}$.

- Find the time step $\boldsymbol{t}$ that minimizes the $\boldsymbol{p}$ -norm distance: $t^* = \arg\min_t \| \sqrt{\beta_t} - \sigma \|_p$



Since $\boldsymbol{t}$ is discrete (1,…,T) → the problem becomes one of searching for the best corresponding state.

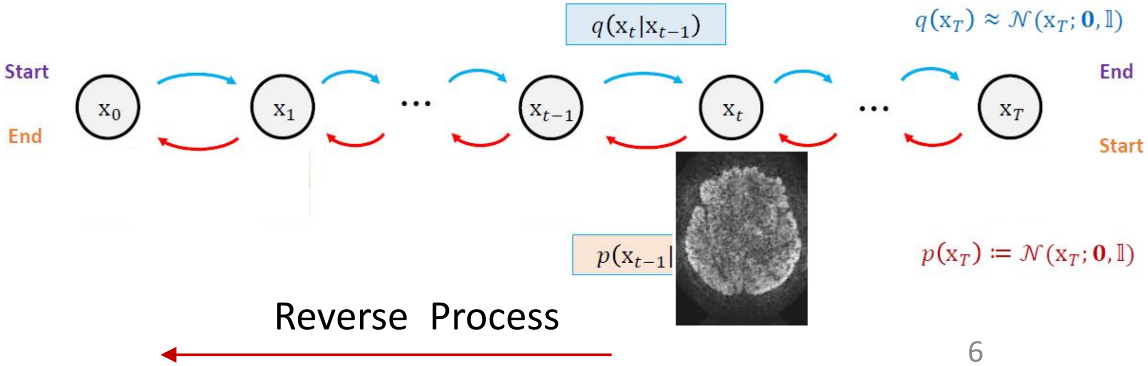
→ Once $\boldsymbol{\beta_t}$ is identified ;

- The **reverse denoising process** can be started from $\mathbf{S_t}$ in the Markov chain.

- The final denoised image S_0 is then progressively obtained via this process ;

$$p(S_0 \mid S_t)$$

$$q(x_t \mid x_{t-1}) = \sqrt{1-\beta_t}\, x_{t-1} + \sqrt{\beta_t}\, z, \quad z \sim \mathcal{N}(0, I).$$



DDM^2 : Stage III

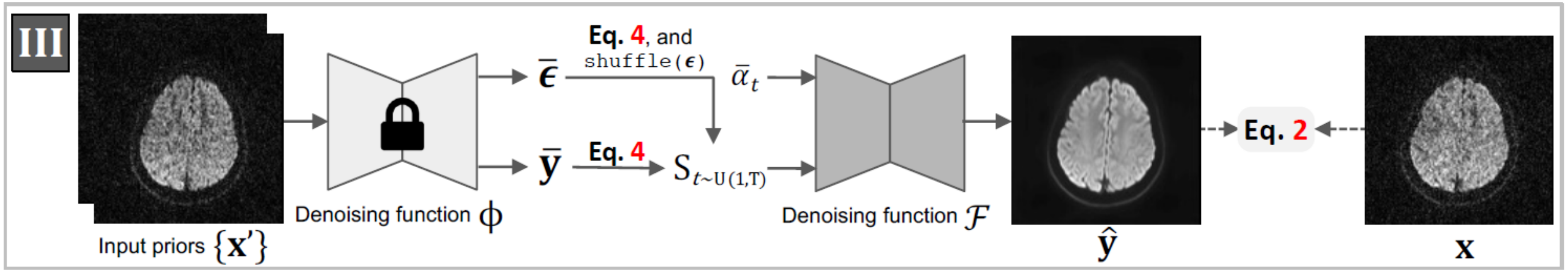
Diffusion Model Training

❖ Train the **generative diffusion model F** to produce clean images ;

Problem :

- If F is trained directly on $\bar{y}$ → **F** may collapse into simply **reproducing Φ**.

→ Leads to suboptimal performance.



Solution :

a) **Noise Shuffle** :

$$q(S_t \mid \bar{y}) = \sqrt{\bar{\alpha}_t}\,\bar{y} + \mathrm{shuffle}(\bar{\varepsilon})$$;

shuffle(·) = spatial shuffling of the residual noise ε estimated by Φ.

- Forces F to learn the true implicit noise distribution of Φ.
- Reduces the gap with the explicit noise model G.

$$\min_F \|F(S_t, \bar{\alpha}_t) - x\|^2$$

b) **J-Invariance Optimization** (loss toward x);

- Instead of constraining F to copy $\bar{y}$(Φ) ; reconstruct an image that reverts back to x after noise is re-added.;

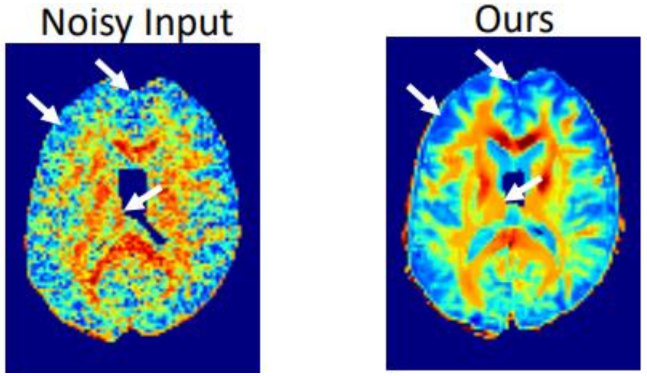
- Corrects approximation errors from Stage I.
- Allows F to explore a broader and more accurate solution space.

$$\bar{\alpha}_t = \prod_{i=1}^{t} \alpha_i, \quad \alpha_t = 1 - \beta_t$$

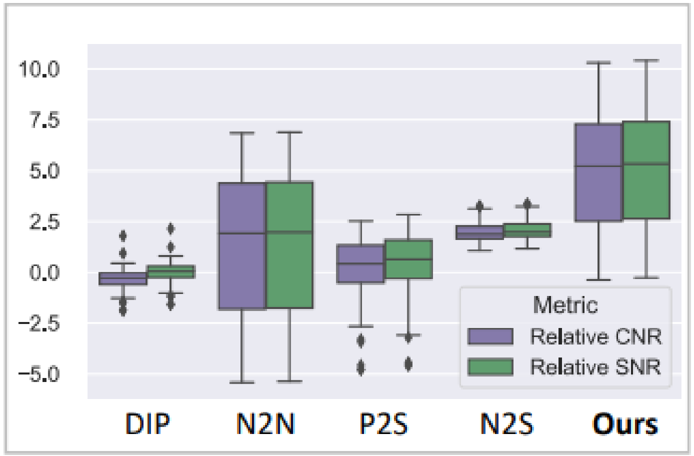
Experiments and Results

- **Qualitative Results :**
  - DDM² removes noise **without over-smoothing** anatomical structures.
  - No hallucinated lesions, as confirmed by two neuroradiologists.

Dataset	Dims (X×Y×Z)	Dirs	Res.	b-values	Reference
gSlider (in-house)	128×128×160	50	0.5 mm iso	1000	In-house
Sherbrooke 3-Shell	128×128×64	193 (eval at b=1000)	–	1000, 2000, 3000	Garyfallidis et al. (2014)
Stanford HARDI	106×81×76	150	–	2000	Rokem (2016)
PPMI (Parkinson)	116×116×72	64	–	2000	Marek et al. (2011)



- **Quantitative Results :**
  - SNR/CNR higher than all competing methods
  - **Average improvement** : +0.95 SNR / +0.93 CNR par compared to N2S

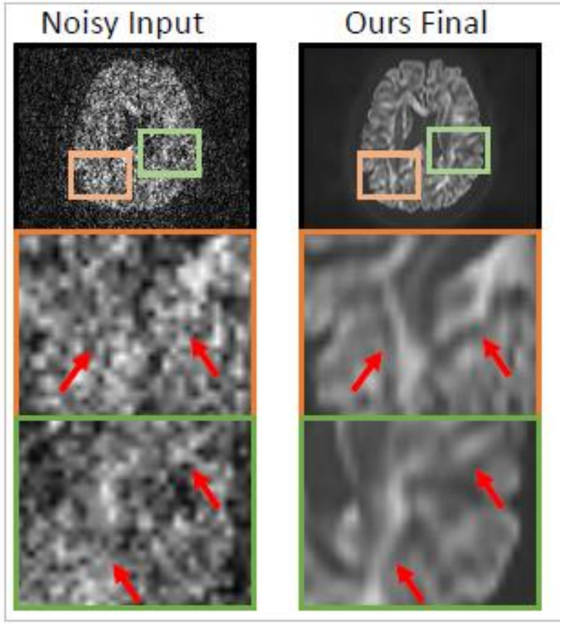


$$\mathrm{SNR} = \frac{\mu_{\text{signal}}}{\sigma_{\text{bruit}}} \qquad \mathrm{CNR} = \frac{|\mu_{\text{foreground}} - \mu_{\text{background}}|}{\sigma_{\text{bruit}}}$$

Foreground := corpus callosum

- DIP (Deep Image Prior)
- Noise2Noise (N2N)
- Patch2Self (P2S)
- Noise2Score (N2S)

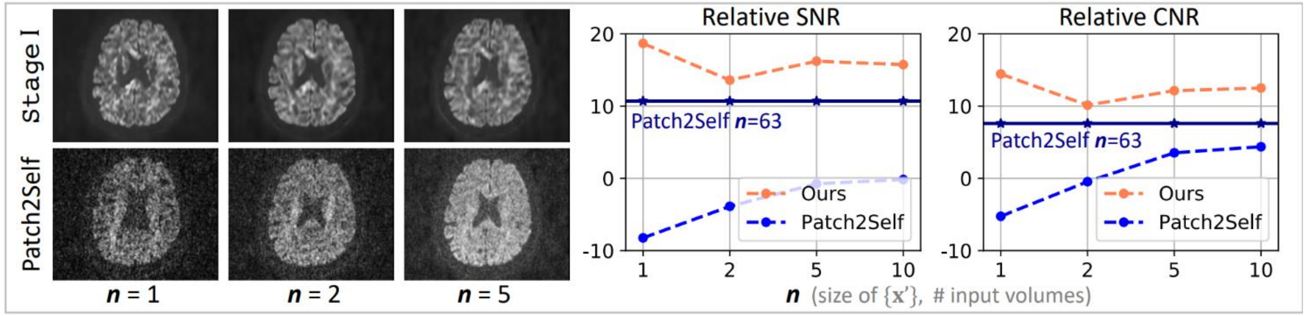
- **Key Advantages :**
  - DDM² = better detail preservation than Patch2Self
  - 5 to 10× **faster than** standard DDPM (thanks to Markov State Matching).



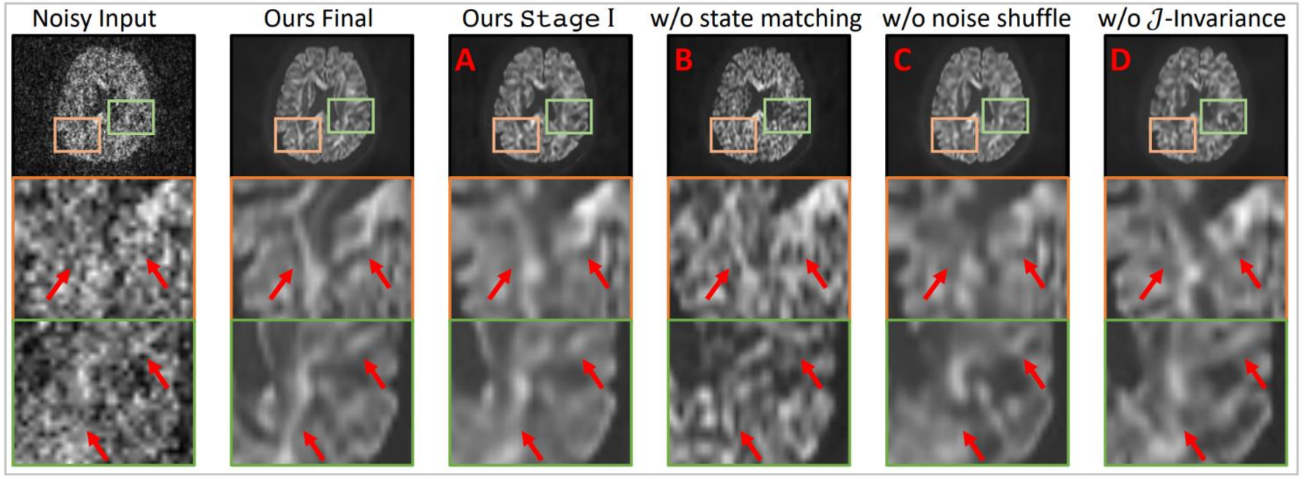
ABLATION STUDIES

Vary number of input volumes

- **Stage I :**
  - Produces clean images even with **n = 1 or 2**
  - Maintains **high SNR/CNR** across all settings
- **Patch2Self :**
  - Fails when **n is small** (images remain noisy)
  - Needs **large n (~63)** to reach acceptable performance



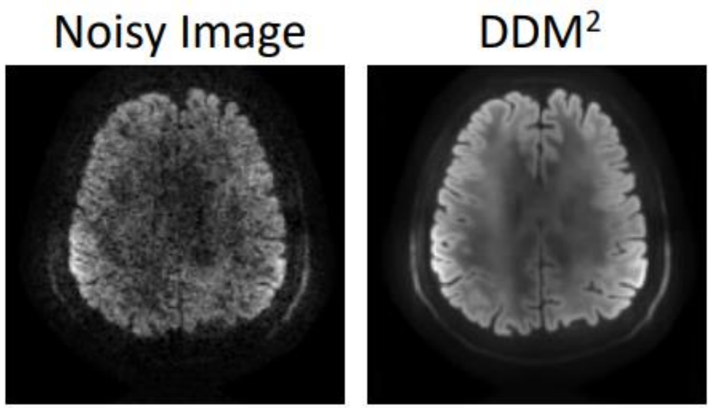
- **Stage I (A)** → good denoising but **too smooth**, loss of fine details
- **w/o State Matching (B)** → **hallucinations**, false anatomical details
- **w/o Noise Shuffle (C)** → structural degradation, loss of robustness
- **w/o J-Invariance (D)** → blurrier results, accumulated errors



Conclusion et Discussion

Summary of the article : DDM2 : Self-supervised denoising in 3 stages

- Noise estimation (Stage I)
- Matching to a Markov state (Stage II)
- Denoised image generation via diffusion (Stage III)



- Achieves higher SNR/CNR with preserved anatomical details.
- Inference time is longer compared to conventional CNNs/MLPs.

Thanks for your attention !