Finanziato dall'Unione europea
NextGenerationEU

Ministero dell'Università e della Ricerca

Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA

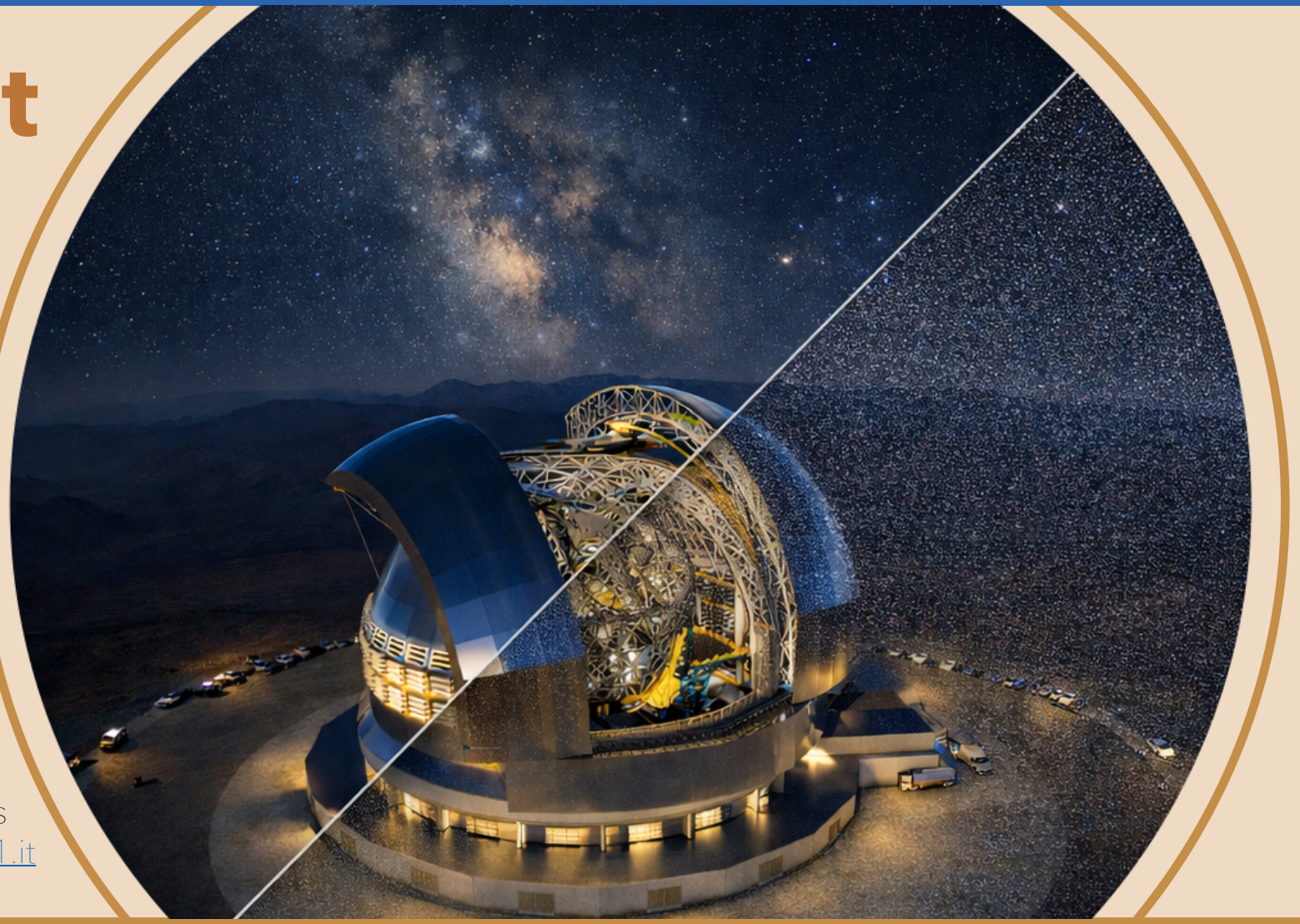
INAF
ISTITUTO NAZIONALE DI ASTROFISICA

Stars do not make a noise

...to be silent is to be beautiful

– James Stephens

A ~~2 hours~~ 40min! :) journey from the classics to the deep learning era of image denoising (...and back!)



Pierpaolo Brutti | Department of Statistics
pierpaolo.brutti@uniroma1.it

PNRR
Missione 4 • Componente 2
Investimento 3.1

STILES – IR0000034
CUP C33C22000640006

Finanziato dall'Unione europea
NextGenerationEU

Ministero dell'Università e della Ricerca

Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA

INAF
ISTITUTO NAZIONALE DI ASTROFISICA



Pierpaolo Brutti
Department of Statistics Science
Sapienza University of Rome

PNRR
Missione 4 • Componente 2
Investimento 3.1

STILES – IR0000034
CUP C33C22000640006

Finanziato dall'Unione europea
NextGenerationEU

Ministero dell'Università e della Ricerca

Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA

INAF
ISTITUTO NAZIONALE DI ASTROFISICA



Spot the intruder

Not exactly a navigated astronomer

PNRR
Missione 4 • Componente 2
Investimento 3.1

STILES – IR0000034
CUP C33C22000640006

Finanziato dall'Unione europea
NextGenerationEU

Ministero dell'Università e della Ricerca

Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA

INAF
ISTITUTO NAZIONALE DI ASTROFISICA



Spot the intruder

Not exactly a navigated astronomer

PNRR
Missione 4 • Componente 2
Investimento 3.1

STILES – IR0000034
CUP C33C22000640006

Finanziato dall'Unione europea
NextGenerationEU

Ministero dell'Università e della Ricerca

Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA

INAF
ISTITUTO NAZIONALE DI ASTROFISICA

...clearly, I'm no Ed Hubble...



PNRR
Missione 4 • Componente 2
Investimento 3.1

STILES – IR0000034
CUP C33C22000640006

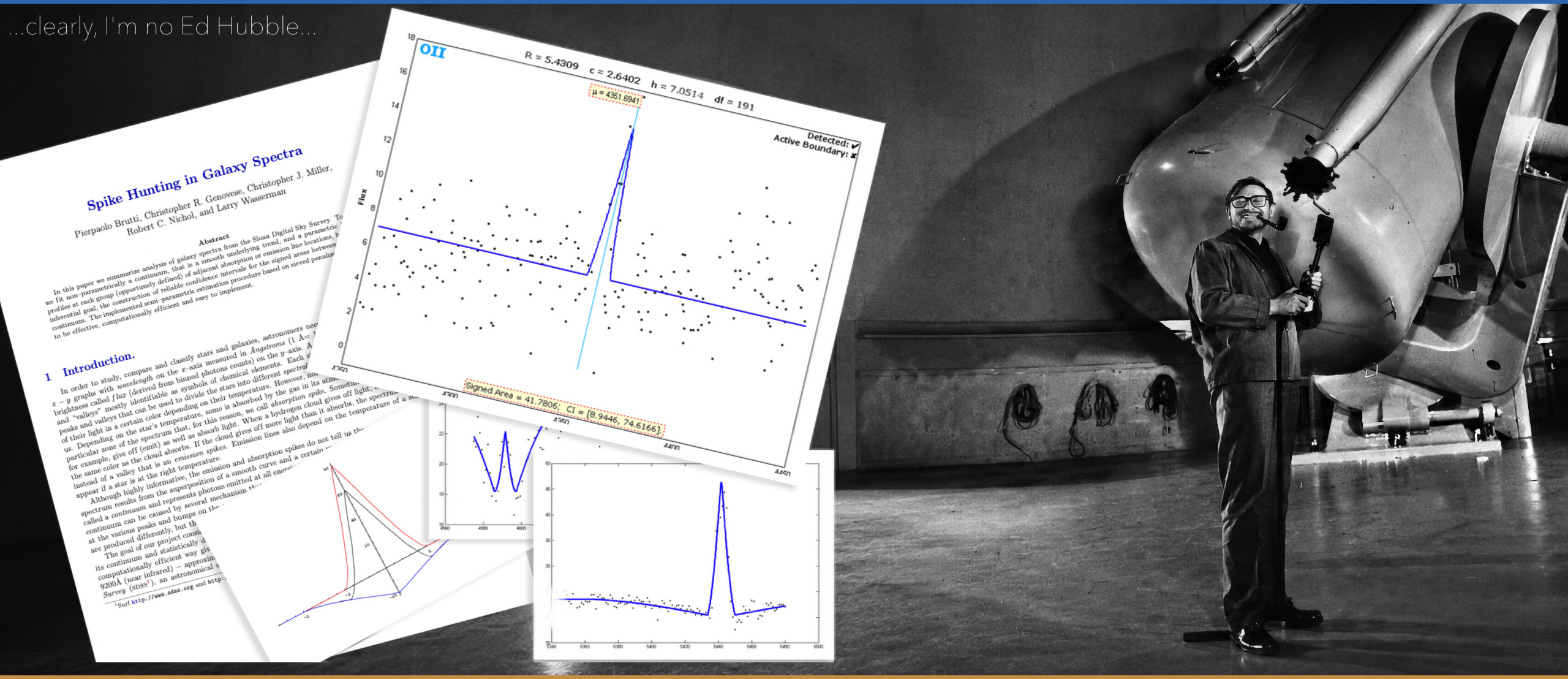
Finanziato dall'Unione europea
NextGenerationEU

Ministero dell'Università e della Ricerca

Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA

INAF
ISTITUTO NAZIONALE DI ASTROFISICA

...clearly, I'm no Ed Hubble...



PNRR
Missione 4 • Componente 2
Investimento 3.1

STILES – IR0000034
CUP C33C22000640006

Finanziato dall'Unione europea
NextGenerationEU

Ministero dell'Università e della Ricerca

Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA

INAF
ISTITUTO NAZIONALE DI ASTROFISICA

OAR@Monte Porzio Catone



Finanziato dall'Unione europea
NextGenerationEU

Ministero dell'Università e della Ricerca

Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA

INAF
ISTITUTO NAZIONALE DI ASTROFISICA

OAR@Monte Porzio Catone











Finanziato dall'Unione europea
NextGenerationEU

Ministero dell'Università e della Ricerca

Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA

INAF
ISTITUTO NAZIONALE DI ASTROFISICA

OAR@Monte Porzio Catone



Astronomy & Astrophysics

A&A, 677, A102 (2023)
https://doi.org/10.1051/0004-6361/202346725
© The Authors 2023

FORECAST: A flexible software to forward model cosmological hydrodynamical simulations mimicking real observations

Flaminia Fortuni[1], Emiliano Merlin[1], Adriano Fontana[1], Carlo Giocoli[1,4], Erik Rossetti[2], Luca Graziani[3], Paola Santini[1], Marco Castellano[1], Stéphane Charlot[5], and Iacopo Chevallard[7]

[1] INAF – Osservatorio Astronomico di Roma, via Frascati 33, 00078 Monte Porzio Catone (Roma), Italy
e-mail: flaminia.fortuni@inaf.it
[2] INAF – Osservatorio Astronomico di Trieste, Via Tiepolo 11, 34131 Trieste, Italy
[3] INAF – Osservatorio di Bologna, Viale Berti Pichat 6/2, 40127 Bologna, Italy
[4] INFN – Sezione di Bologna, Viale Berti Pichat 6/2, 40127 Bologna, Italy
[5] INAF – Osservatorio di Astrofisica e Scienza dello Spazio di Bologna, via Gobetti 93/3, 40129 Bologna, Italy
[6] Dipartimento di Fisica, Università di Roma "La Sapienza", Piazzale Aldo Moro 5, 00185 Roma, Italy
[7] Sorbonne Université, CNRS, UMR 7095, Institut d'Astrophysique de Paris, 98 bis bd Arago, 75014 Paris, France
[8] Department of Physics, University of Oxford, Denys Wilkinson Building, Keble Road, Oxford OX1 3RH, UK

Received 21 April 2023 / Accepted 18 July 2023

ABSTRACT

Context. Comparing theoretical predictions to real data is crucial to properly formulate galaxy formation and evolution models. Historically done entirely considering the direct output of simulations and quantities inferred from observations, leading to inconsistencies.

Aims. We present FORECAST, a new flexible and adaptable software package that performs forward modelling of synthetic cosmological hydrodynamical simulations to create a wide range of realistic synthetic surveys. It provides fundamental tools for accurate comparisons with observational data. With customizable options, FORECAST constructs a light cone centered on the observer's position exploiting the snapshots of a simulation. Its relevant parameters, it allows users to tailor the synthetic images to their specific requirements.

Methods. FORECAST constructs a light cone centered on the observer's position exploiting the snapshots of a simulation. Its computes the observed flux of each simulated stellar element, modeled as a single stellar population, including intergalactic medium absorption, and dust attenuation. Finally, it convolves the flux with realistic morphologies and size evolution, and adds instrumental and observational features such as background noise and PSF effects, including k corrections, to which observational features such as background noise and PSF effects, including k corrections, to which observational features such as background noise and PSF effects, including k corrections, to which

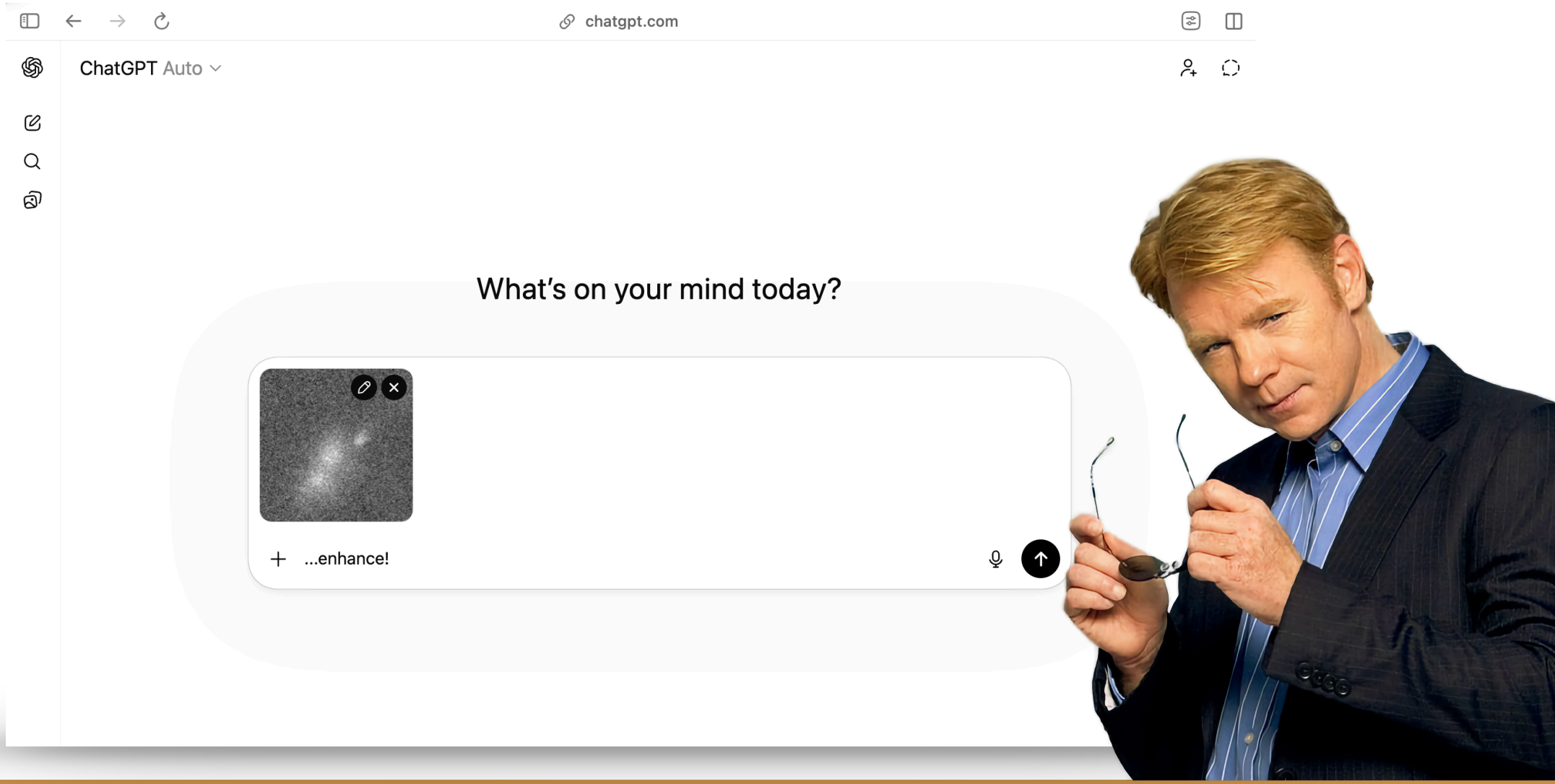
. . . it's 2026, so . . .

Finanziato dall'Unione europea
NextGenerationEU

Ministero dell'Università e della Ricerca

Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA

INAF
ISTITUTO NAZIONALE DI ASTROFISICA



Finanziato dall'Unione europea
NextGenerationEU

Ministero dell'Università e della Ricerca

Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA

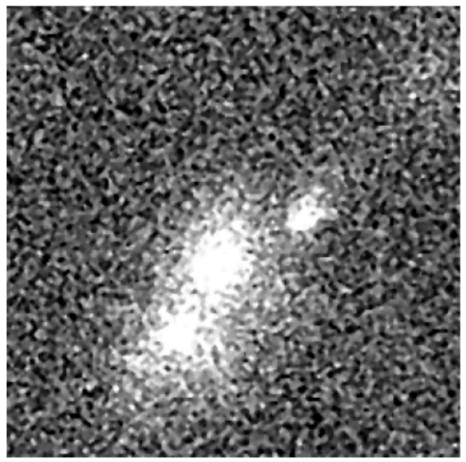
INAF
ISTITUTO NAZIONALE DI ASTROFISICA



https://tinyurl.com/27e7ffdr

chatgpt.com

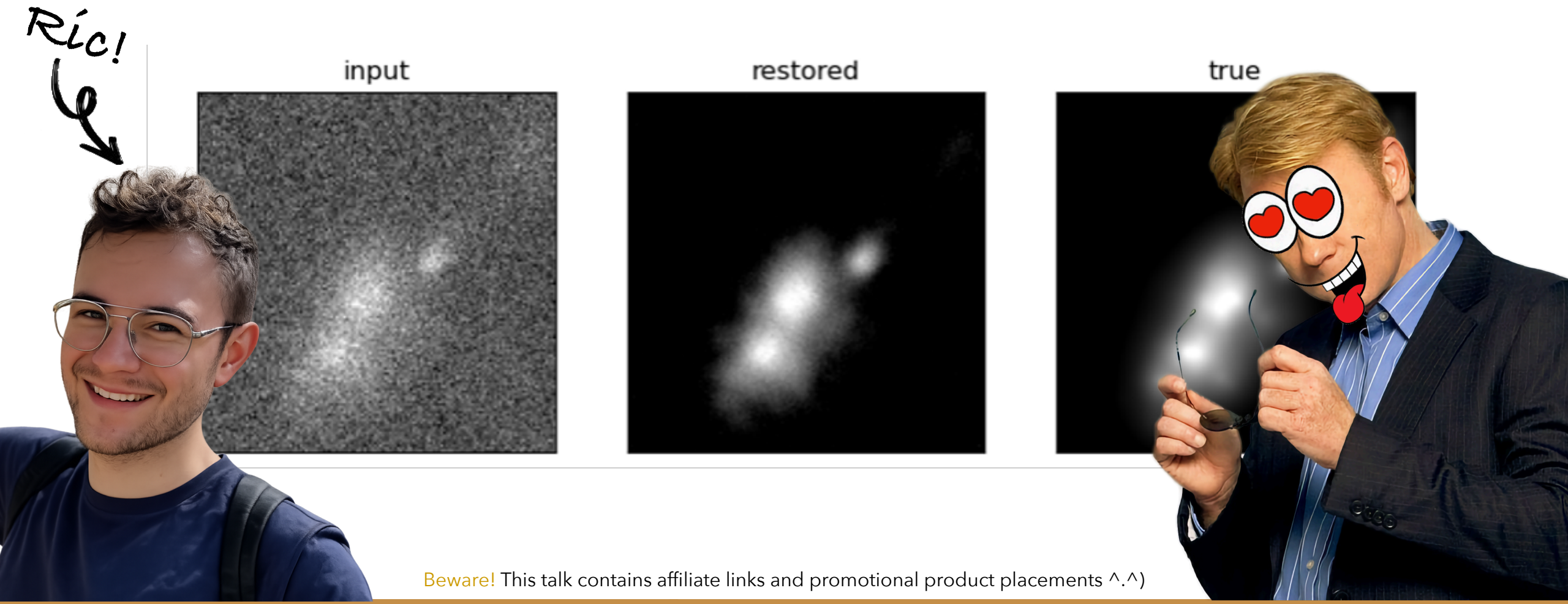
ChatGPT Auto



I enhanced the image by reducing noise, increasing contrast, and sharpening the bright structures so the central features stand out more clearly.







Beware! This talk contains affiliate links and promotional product placements ^.^)

Finanziato dall'Unione europea
NextGenerationEU

Ministero dell'Università e della Ricerca

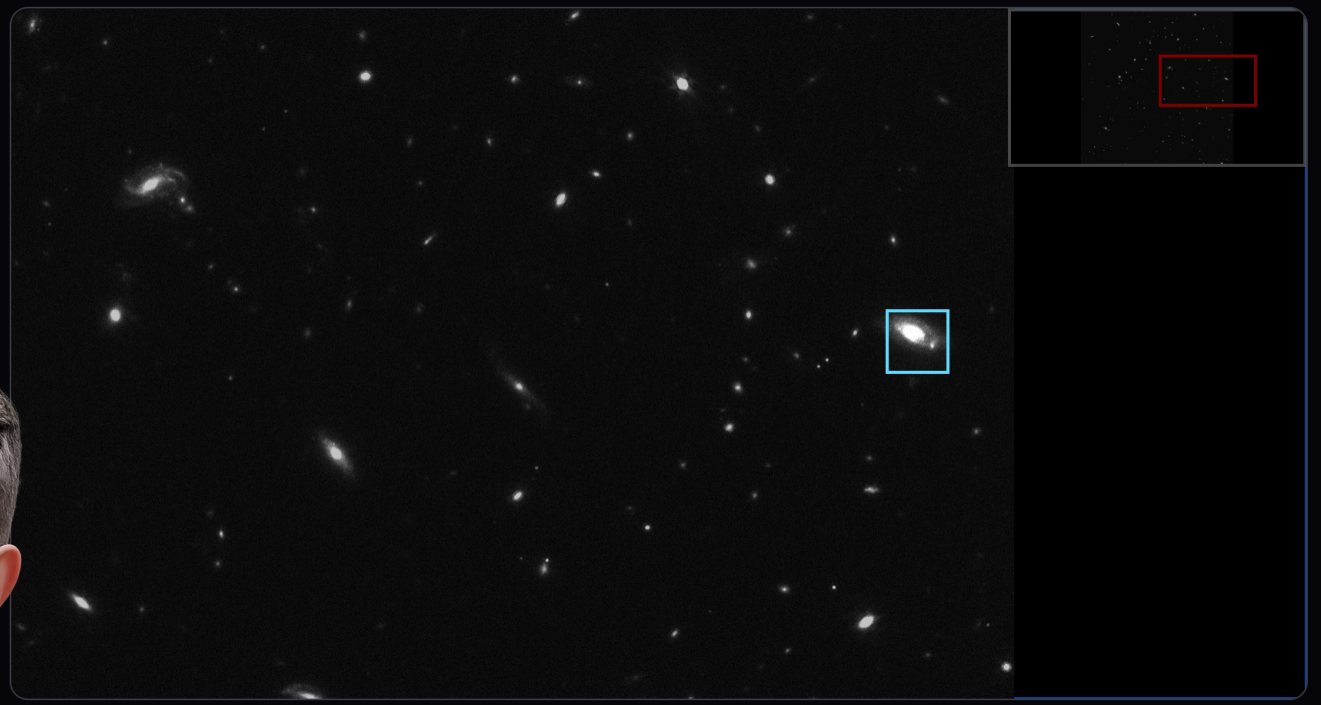
Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA

INAF
ISTITUTO NAZIONALE DI ASTROFISICA



Astro Denoising (FORECAST)

linear | zscale



x=3853.9 y=1766.9 | patch=(3840, 1760) | pid=6995

Method	MSE	PSNR	SSIM
[illegible]	1.687e-6	39.44	0.9530
[illegible]	4.044e-6	35.64	0.9180
[illegible]	5.928e-6	33.98	0.7347

Patch 128x128



Finanziato dall'Unione europea
NextGenerationEU

Ministero dell'Università e della Ricerca

Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA

INAF
ISTITUTO NAZIONALE DI ASTROFISICA

Astro Denoising (FORECAST)

linear zscale

Patch 128x128

IMAGE CLEAN AUTOENCODER BM3D

DDPM GAUSSIAN MEDIAN NL_MEANS

ONAMALIK PERSTREE TOPOATT UNET

WAVELET

x=3853.9 y=1766.9 | patch=(3840, 1760) | pid=6995

Method	MSE	PSNR	SSIM
	1.687e-6	39.44	0.9530
	4.044e-6	35.64	0.9130
	5.928e-6	33.98	0.7347

Ric!



https://tinyurl.com/2puyc73y

. . . let's cut to the chase . . .

Main source: Elad et al. (2023). Image Denoising: The Deep Learning Revolution and Beyond – A Survey Paper

Today

a short list of good intentions

> Grand Goal
>
> Take this chance to know your community better

1. A brief recap of what happened in the field of denoising in the last…uhmm, three decades (overreaching a bit here ^.^')
2. A first pointer to new developments I personally enjoy: diffusion models
3. An optional (i.e. opt-in, you tell me!) deeper dive into diffusion models
4. Depending on time, one or two specific applications of recent proposals

Finanziato dall'Unione europea
NextGenerationEU

Ministero dell'Università e della Ricerca

Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA

INAF
ISTITUTO NAZIONALE DI ASTROFISICA



Disclaimer:
I'll make tons of approximations, be patient, thx!

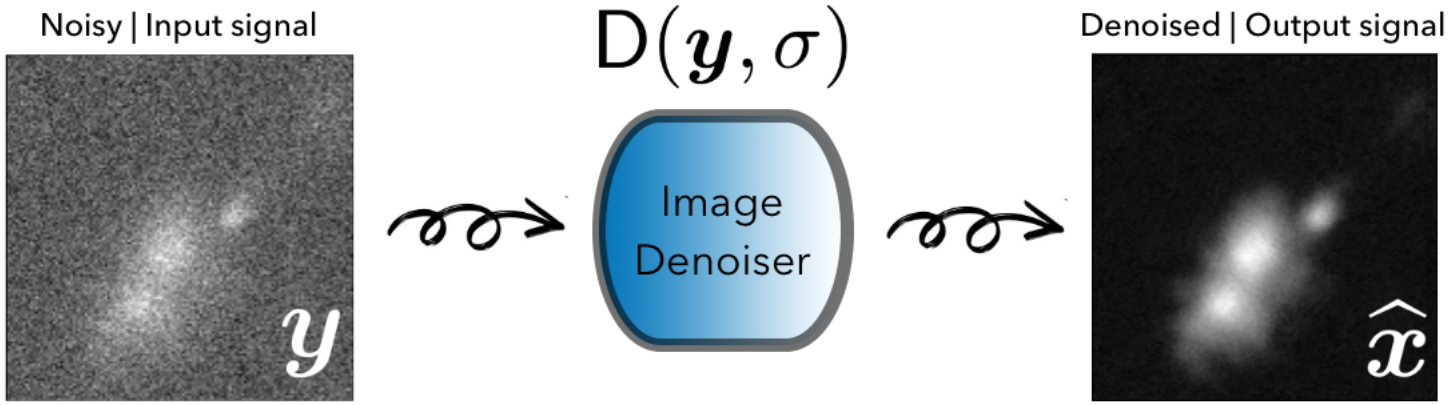


Image Denoising

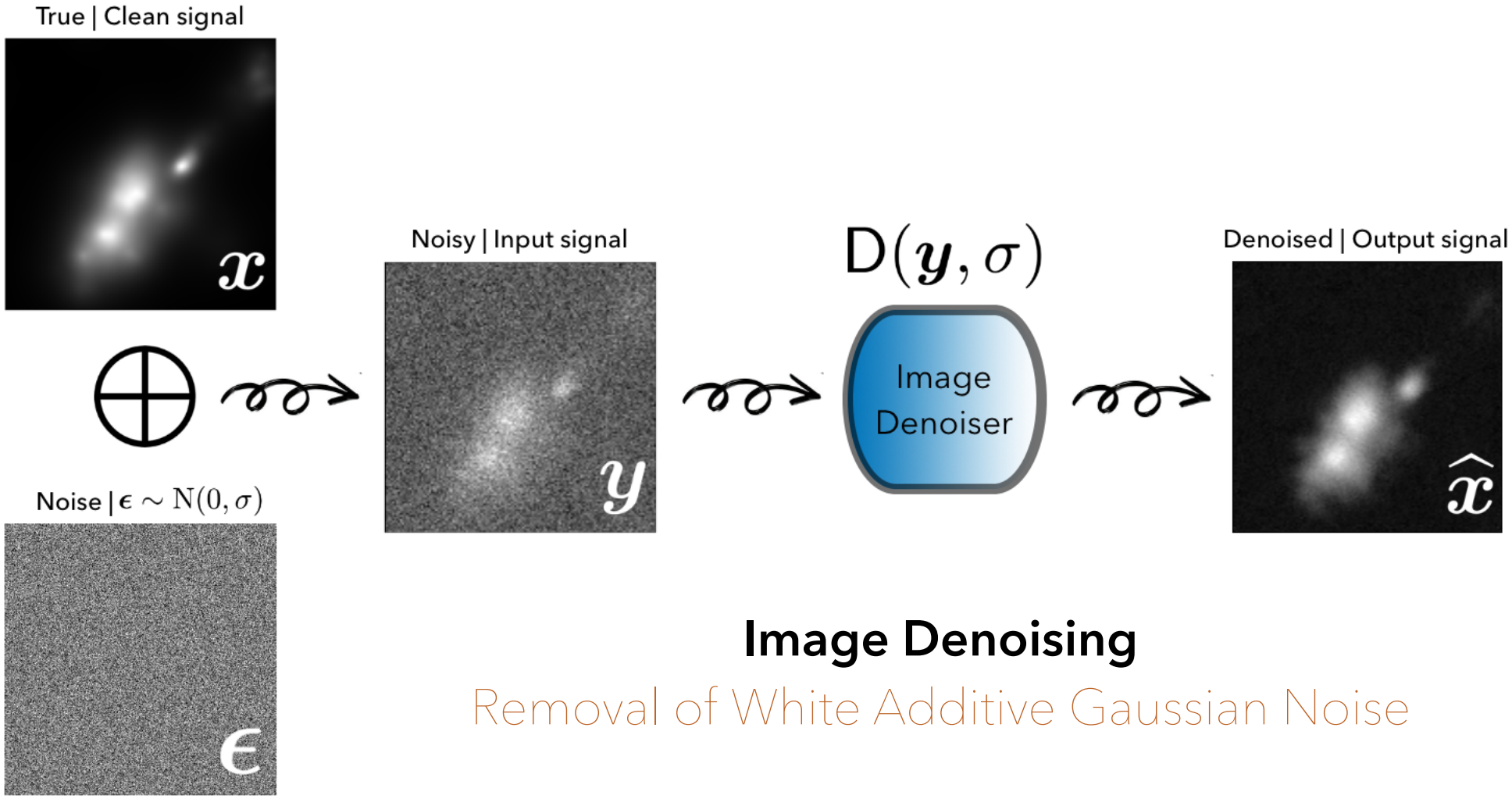


Image Denoising

Removal of White Additive Gaussian Noise

- **Poisson noise** useful but well approximated by Gaussian in high/low photon count scenarios, possibly Anscombe's variance-stabilizing transformation.

- Many developments under gaussianity can be converted to other noise models, also of mixed type (Poisson-Gaussian)

- Optimal denoisers under gaussianity play a key role in modern **generative models**!

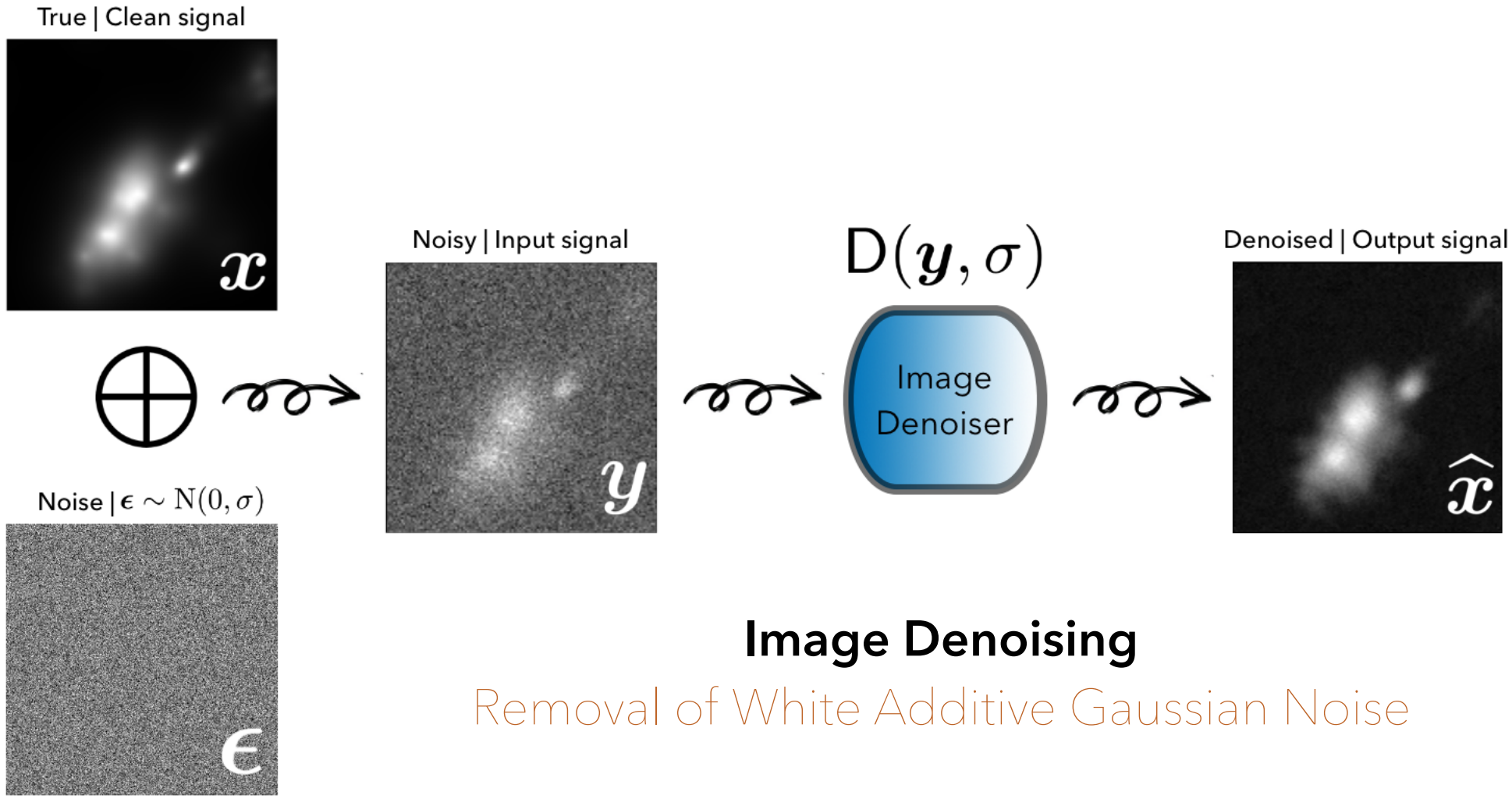


Image Denoising

Removal of White Additive Gaussian Noise

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
INAF ISTITUTO NAZIONALE DI ASTROFISICA

Recap of the basics

Observation Model

$$\boldsymbol{y} = \boldsymbol{x} + \boldsymbol{\epsilon} \quad \boldsymbol{\epsilon} \sim \mathrm{N}(\mathbf{0}, \sigma^2 \mathbb{I})$$

Ideal: MSE-Optimal denoiser (MMSE)

$$\widehat{\boldsymbol{x}}_{\mathtt{MMSE}} = \underset{\mathtt{All}\ \widehat{\boldsymbol{x}}}{\operatorname{argmin}}\, \mathbb{E}\left(\|\boldsymbol{x} - \widehat{\boldsymbol{x}}\|^2\right) = \mathbb{E}(\boldsymbol{x}|\boldsymbol{y}) = \int \boldsymbol{x}\, \underset{\mathtt{unknown}}{p(\boldsymbol{x}|\boldsymbol{y})}\, \mathrm{d}\boldsymbol{x}.$$



Likelihood:

$$p(\boldsymbol{y}|\boldsymbol{x}) \propto \exp\left\{-\frac{\|\boldsymbol{y} - \boldsymbol{x}\|^2}{2\sigma^2}\right\}$$

Posterior:

$$p(\boldsymbol{x}|\boldsymbol{y}) \propto p(\boldsymbol{y}|\boldsymbol{x})\, p(\boldsymbol{x})$$

where $p(\boldsymbol{x})$ is the **image prior**

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
INAF ISTITUTO NAZIONALE DI ASTROFISICA

Recap of the basics

Observation Model

$$\boldsymbol{y} = \boldsymbol{x} + \boldsymbol{\epsilon} \quad \boldsymbol{\epsilon} \sim \mathrm{N}(\mathbf{0}, \sigma^2 \mathbb{I})$$

Ideal: MSE-Optimal denoiser (MMSE)

$$\widehat{\boldsymbol{x}}_{\texttt{MMSE}} = \underset{\texttt{All } \widehat{\boldsymbol{x}}}{\operatorname{argmin}} \, \mathbb{E}\left(\|\boldsymbol{x} - \widehat{\boldsymbol{x}}\|^2 \right) = \mathbb{E}(\boldsymbol{x}|\boldsymbol{y}) = \int \boldsymbol{x} \, \underset{\texttt{unknown}}{p(\boldsymbol{x}|\boldsymbol{y})} \, \mathrm{d}\boldsymbol{x}.$$



Maximum Likelihood:

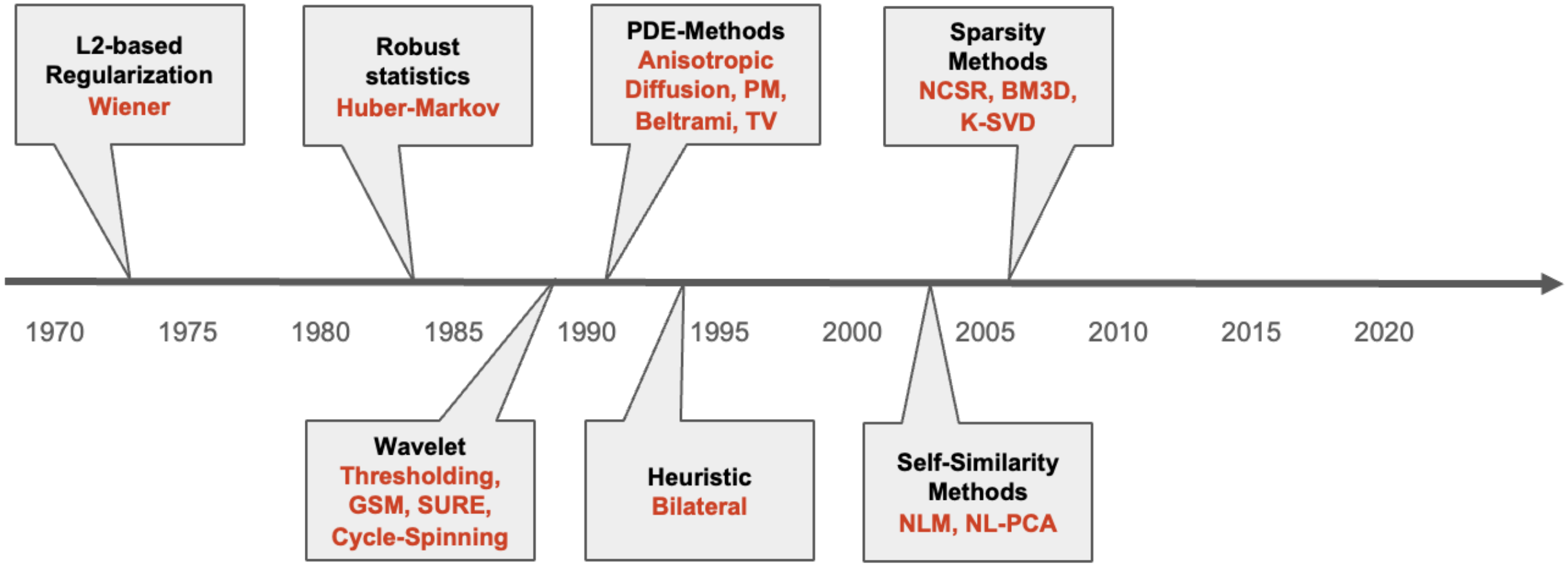
$$\widehat{\boldsymbol{x}}_{\texttt{MLE}} = \underset{\texttt{All } \boldsymbol{x}}{\operatorname{argmax}} \, p(\boldsymbol{y}|\boldsymbol{x}) = \boldsymbol{y}$$

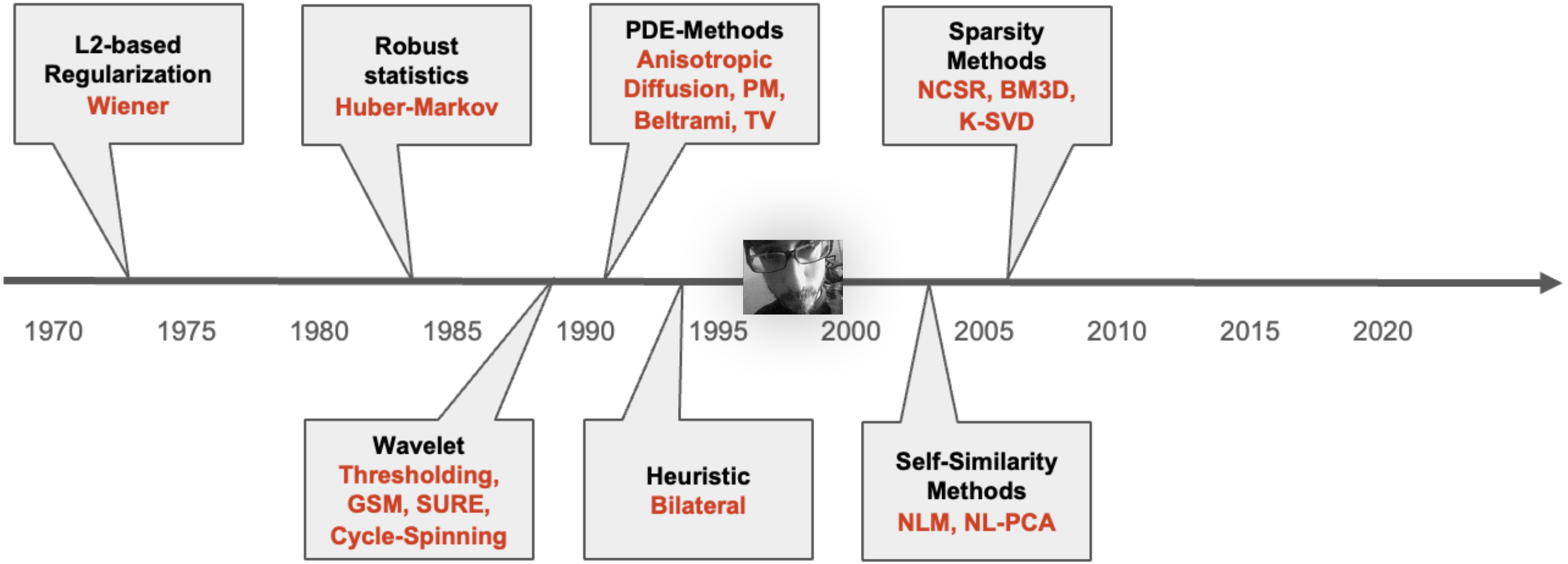
No denoising $\rightsquigarrow$ we need a prior.

Maximum a Posterior:

$$\widehat{\boldsymbol{x}}_{\texttt{MAP}} = \underset{\texttt{All } \boldsymbol{x}}{\operatorname{argmax}} \, p(\boldsymbol{x}|\boldsymbol{y})$$

$$= \underset{\texttt{All } \boldsymbol{x}}{\operatorname{argmin}} \, \frac{1}{2}\|\boldsymbol{y} - \boldsymbol{x}\|^2 - \log p(\boldsymbol{x})$$


L2-based Regularization
Wiener
Robust statistics
Huber-Markov
PDE-Methods
Anisotropic Diffusion, PM, Beltrami, TV
Sparsity Methods
NCSR, BM3D, K-SVD
Wavelet
Thresholding, GSM, SURE, Cycle-Spinning
Heuristic
Bilateral
Self-Similarity Methods
NLM, NL-PCA
1970
1975
1980
1985
1990
1995
2000
2005
2010
2015
2020


L2-based Regularization
Wiener
Robust statistics
Huber-Markov
PDE-Methods
Anisotropic Diffusion, PM, Beltrami, TV
Sparsity Methods
NCSR, BM3D, K-SVD
1970
1975
1980
1985
1990
1995
2000
2005
2010
2015
2020
Wavelet
Thresholding, GSM, SURE, Cycle-Spinning
Heuristic
Bilateral
Self-Similarity Methods
NLM, NL-PCA



L2-based Regularization
Wiener

Robust statistics
Huber-Markov

PDE-Methods
Anisotropic Diffusion, PM, Beltrami, TV

Spars Meth
NCSR, K-SV

David L. DONOHO and Iain M. JOHNSTONE

Adapting to Unknown Smoothness via Wavelet Shrinkage



1970 1975 1980 1985 1990 1995 2000 2005 2010 2015 2020

Wavelet
Thresholding, GSM, SURE, Cycle-Spinning

Heuristic
Bilateral

Self-Similarity Methods
NLM, NL-PCA



L2-based Regularization
Wiener

Robust statistics
Huber-Markov

PDE-Methods
Anisotropic Diffusion, PM, Beltrami, TV

Spars
Meth
NCSR,
K-SV



Wavelet
Thresholding, GSM, SURE, Cycle-Spinning

Heuristic
Bilateral

Self-Simil
N

1970 1975 1980 1985 1990 1995 2000 2005 2010

David L. DONOHO and Iain M. JOHNSTONE

Adapting to Unknown Smoothness via Wavelet Shrinkage



Mon. Not. R. Astron. Soc. **383**, 539–545 (2008)

Spherical needlets for cosmic microwave background data analysis

D. Marinucci,[1⋆] D. Pietrobon,[2] A. Balbi,[2,3] P. Baldi,[1] P. Cabella,[4] G. Kerkyacharian,[5]
P. Natoli,[2,3] D. Picard[6] and N. Vittorio[2,3]



L2-based Regularization
Wiener

Robust statistics
Huber-Markov

PDE-Methods
Anisotropic Diffusion, PM, Beltrami, TV

Spars
Meth
NCSR,
K-SV



Wavelet
Thresholding, GSM, SURE, Cycle-Spinning

Heuristic
Bilateral

Self-Simil
N

1970 1975 1980 1985 1990 1995 2000 2005 2010

David L. DONOHO and Iain M. JOHNSTONE

Adapting to Unknown Smoothness via Wavelet Shrinkage



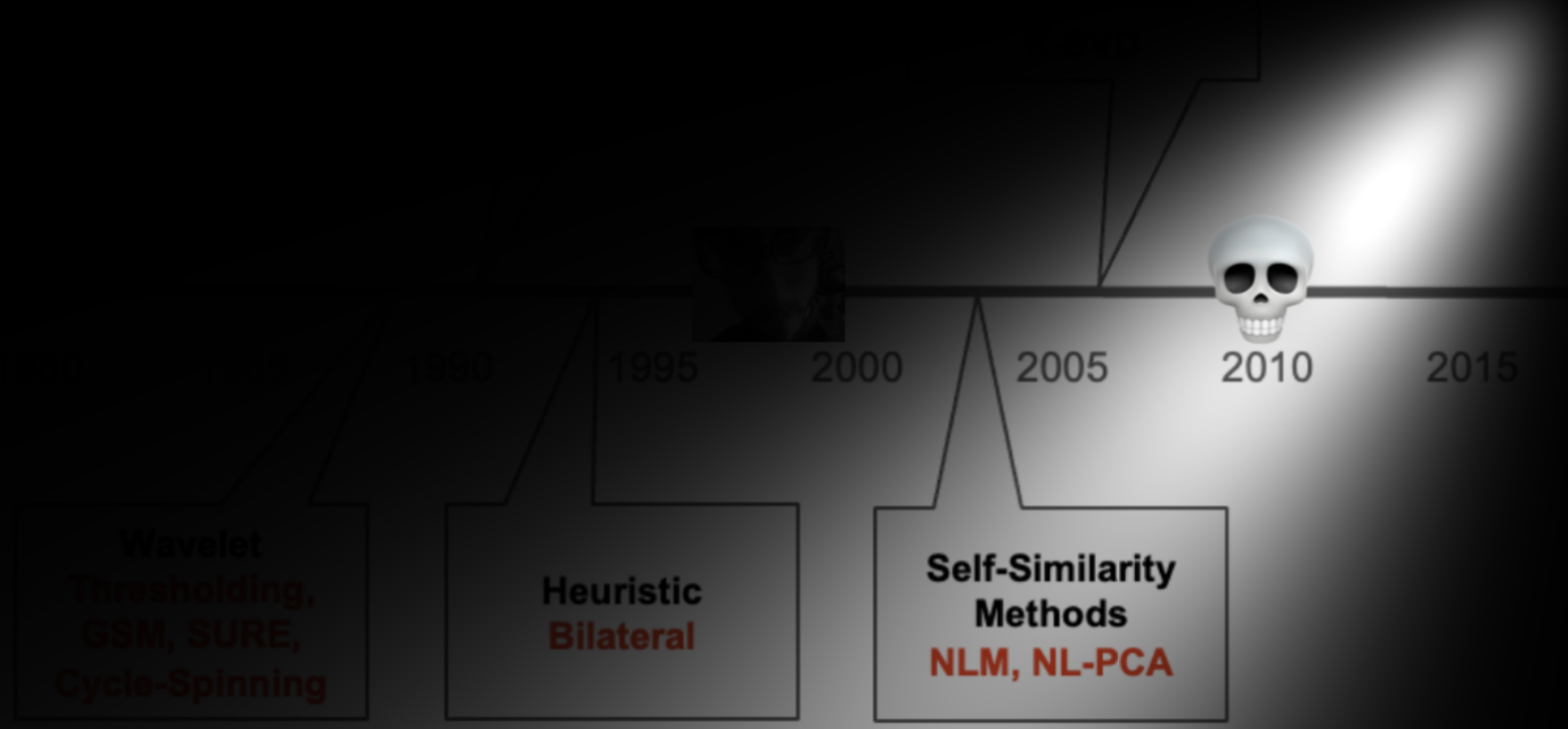
Mon. Not. R. Astron. Soc. 383, 539–545 (2008)

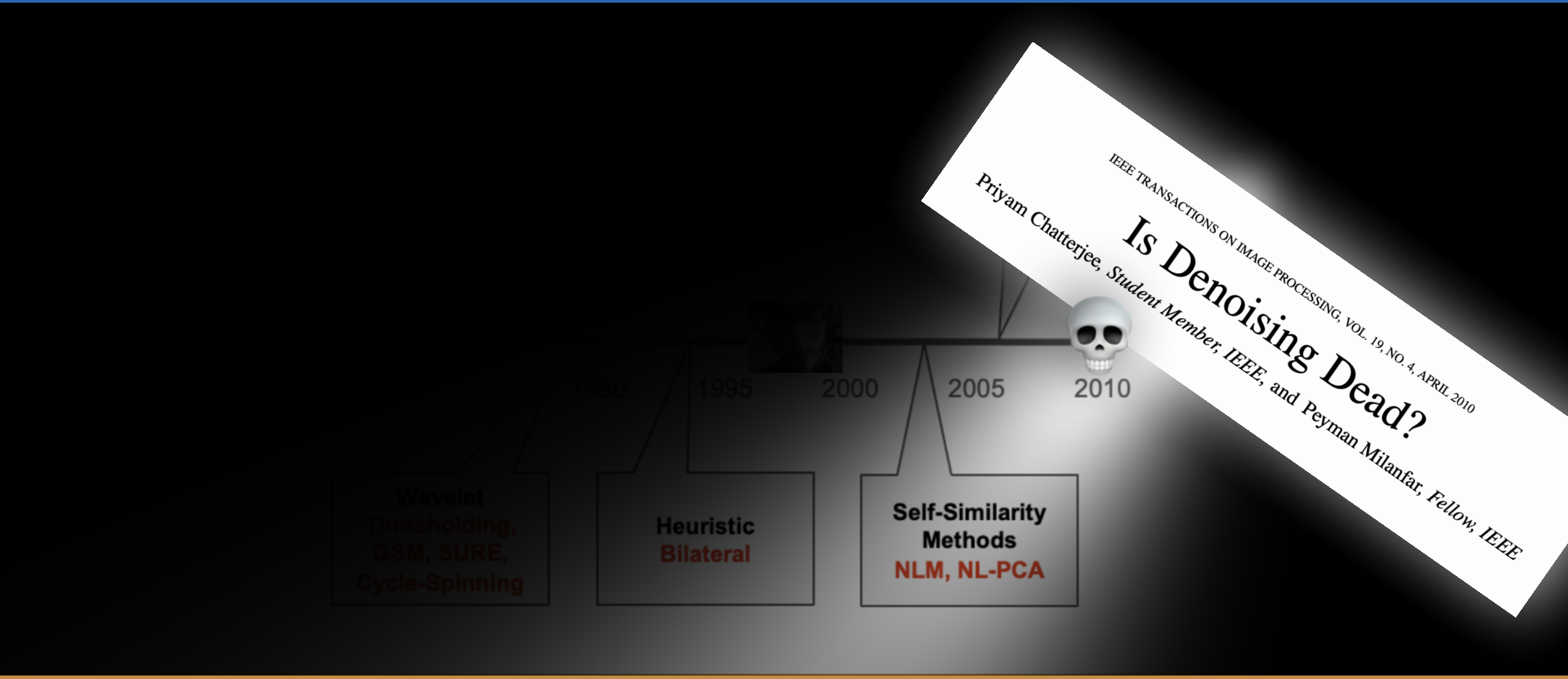
Spherical needlets for cosmic microwave background data analysis

D. Marinucci,[1]★ D. Pietrobon,[2] A. Balbi,[2,3] P. Baldi,[1] P. Cabella,[4] G. Kerkyacharian,[5]
P. Natoli,[2,3] D. Picard[6] and N. Vittorio[2,3]

Finanziato dall'Unione europea
NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA
INAF
ISTITUTO NAZIONALE DI ASTROFISICA

. . . but then . . .


1990
1995
2000
2005
2010
2015
Wavelet Thresholding, GSM, SURE, Cycle-Spinning
Heuristic Bilateral
Self-Similarity Methods NLM, NL-PCA


1990
1995
2000
2005
2010
Wavelet Thresholding, GSM, SURE, Cycle-Spinning
Heuristic Bilateral
Self-Similarity Methods NLM, NL-PCA

IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 19, NO. 4, APRIL 2010
Is Denoising Dead?
Priyam Chatterjee, Student Member, IEEE, and Peyman Milanfar, Fellow, IEEE

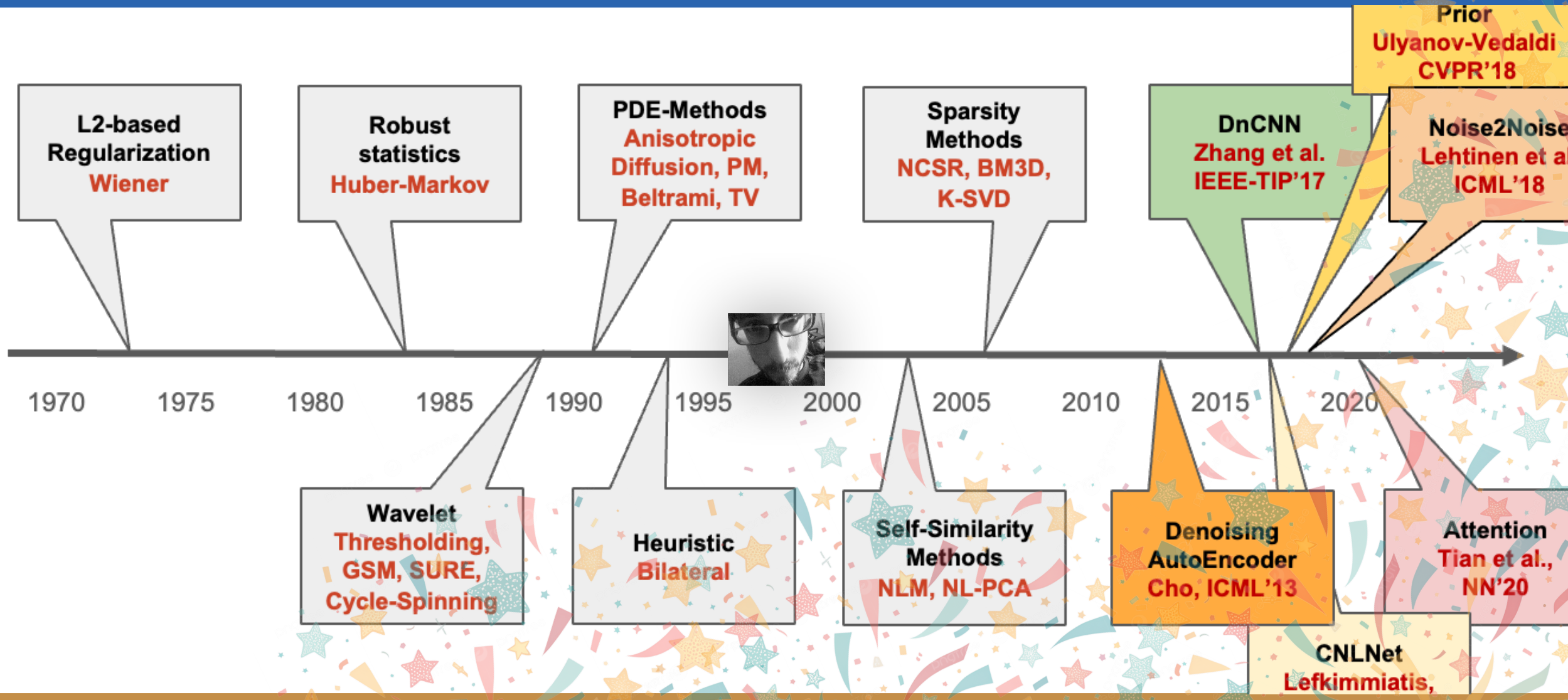
Finanziato dall'Unione europea
NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA
INAF
ISTITUTO NAZIONALE DI ASTROFISICA





. . . but then . . .

Finanziato dall'Unione europea
NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA
INAF
ISTITUTO NAZIONALE DI ASTROFISICA



PNRR
Missione 4 • Componente 2
Investimento 3.1
STILES - IR0000034
CUP C33C22000640006

Classical Principles

- Low dimensionality
- Sparse Representation
- Local Smoothness
- Non-local self-similarity
- Low Rank

Years	Core concept	Formulae for prior
∼1970	Energy regularization	$\|\boldsymbol{x}\|_2^2$
1975–1985	Spatial smoothness	$\|L\boldsymbol{x}\|_2^2$ or $\|D_v\boldsymbol{x}\|_2^2 + \|D_h\boldsymbol{x}\|_2^2$
1980–1985	Optimally learned transform	$\|T\boldsymbol{x}\|_2^2 = \boldsymbol{x}^\mathsf{T} R^{-1}\boldsymbol{x}$ (via PCA)
1980–1990	Weighted smoothness	$\|L\boldsymbol{x}\|_{2,W}^2$
1990–2000	Robust statistics	$\mathbf{1}^\mathsf{T}\mu\{L\boldsymbol{x}\}$ (e.g. Huber–Markov)
1992–2005	Total Variation (TV)	$\int_{v\in\Omega} \lvert\nabla\boldsymbol{x}(v)\rvert\, dv = \mathbf{1}^\mathsf{T}\sqrt{\lvert D_v\boldsymbol{x}\rvert^2 + \lvert D_h\boldsymbol{x}\rvert^2}$
1987–2005	Other PDE-based options	$\int_{v\in\Omega} g\big(\nabla\boldsymbol{x}(v), \nabla^2\boldsymbol{x}(v)\big)\, dv$
2005–2009	Field-of-experts	$\sum_k \lambda_k \mathbf{1}^\mathsf{T}\mu_k\{L_k\boldsymbol{x}\}$
1993–2005	Wavelet sparsity	$\|W\boldsymbol{x}\|_1$
2000–2010	Self-similarity	$\sum_k \sum_{j\in\Omega(k)} d\{R_k\boldsymbol{x}, R_j\boldsymbol{x}\}$
2002–2012	Sparsity methods	$\|\boldsymbol{\alpha}\|_0$ s.t. $\boldsymbol{x} = D\boldsymbol{\alpha}$
2010–2017	Low-rank assumption	$\sum_k \|X_{\Omega(k)}\|_\star$

Classical Principles



"just"…

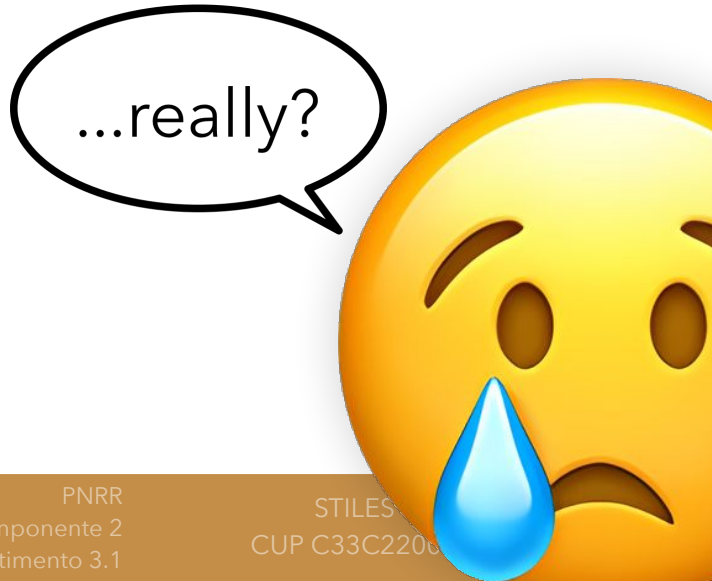
Supervised Training



. . .still, plenty to be happy about!



. . .still, plenty to be happy about!



1st pass

From **denoising** to **sampling** and back...



- At the turn of this decade, the idea of using a good **denoiser** to **synthesize** (*hallucinate*) new images emerged.



- At the turn of this decade, the idea of using a good **denoiser** to **synthesize** (*hallucinate*) new images emerged.

- In essence:
  Use the **denoiser** as a **projector** onto the *image manifold*.



- At the turn of this decade, the idea of using a good **denoiser** to **synthesize** (*hallucinate*) new images emerged.

- In essence:
  Use the **denoiser** as a **projector** onto the *image manifold*.

- Main (backward) steps:
  1. Start with (or *get to*) a **random noisy** image.
  2. "Climb" to a more probable image by **Langevin Dynamics.**
  3. …and look back at the Bayesian literature of the '60s…surprise!



- At the turn of this decade, the idea of using a good **denoiser** to **synthesize** (*hallucinate*) new images emerged.

- In essence:
  Use the **denoiser** as a **projector** onto the *image manifold*.

- Main (backward) steps:
  1. Start with (or *get to*) a **random noisy** image.
  2. "Climb" to a more probable image by **Langevin Dynamics.**
  3. …and look back at the Bayesian literature of the '60s…surprise!

Diffusion Models Beat GANs on Image Synthesis

Prafulla Dhariwal*
OpenAI
prafulla@openai.com

Alex Nichol*
OpenAI
alex@openai.com

Abstract

We show that diffusion models can achieve image sample quality superior to the current state-of-the-art generative models. We achieve this on unconditional image synthesis by finding a better architecture through a series of ablations. For conditional image synthesis, we further improve sample quality with classifier guidance: a simple, compute-efficient method for trading off diversity for fidelity using gradients from a classifier. We achieve an FID of 2.97 on ImageNet 128×128, 4.59 on ImageNet 256×256, and 7.72 on ImageNet 512×512, and we match BigGAN-deep even with as few as 25 forward passes per sample, all while maintaining better coverage of the distribution. Finally, we find that classifier guidance combines well with upsampling diffusion models, further improving FID to 3.94 on ImageNet 256×256 and 3.85 on ImageNet 512×512. We release our code at https://github.com/openai/guided-diffusion.

1 Introduction



Midjourney

Stable Diffusion

DALL-E

Adobe Firefly

*Remark: in these examples the generation is conditional to a prompt.

- At the turn of this decade, the idea of using a good **denoiser** to **synthesize** (*hallucinate*) new images emerged.

- In essence:
  Use the **denoiser** as a **projector** onto the *image manifold*.

- Main (backward) steps:
  1. Start with (or *get to*) a **random noisy** image.
  2. "Climb" to a more probable image by **Langevin Dynamics.**
  3. …and look back at the Bayesian literature of the '60s…surprise!

- Pretend

  We can access the **true/clean** image distribution $p(\boldsymbol{x})$

- Goal

  Draw a sample from $p(\boldsymbol{x})$

- At the turn of this decade, the idea of using a good **denoiser** to **synthesize** (*hallucinate*) new images emerged.

- In essence:
  Use the **denoiser** as a **projector** onto the *image manifold*.

- Main (backward) steps:
  1. Start with (or *get to*) a **random noisy** image.
  2. "Climb" to a more probable image by **Langevin Dynamics.**
  3. …and look back at the Bayesian literature of the '60s…surprise!

- Pretend

  We can access the **true/clean** image distribution $p(\boldsymbol{x})$

- Goal

  Draw a sample from $p(\boldsymbol{x})$

  Step0: start with a purely random noise image $\widehat{\boldsymbol{x}}_0$

- At the turn of this decade, the idea of using a good **denoiser** to **synthesize** (*hallucinate*) new images emerged.

- In essence:
  Use the **denoiser** as a **projector** onto the *image manifold*.

- Main (backward) steps:
  1. Start with (or *get to*) a **random noisy** image.
  2. "Climb" to a more probable image by **Langevin Dynamics.**
  3. …and look back at the Bayesian literature of the '60s…surprise!

- Pretend

  We can access the **true/clean** image distribution $p(\boldsymbol{x})$

- Goal

  Draw a sample from $p(\boldsymbol{x})$

  Step0: start with a purely random noise image $\hat{\boldsymbol{x}}_0$

  Step1: climb to a more probable image by iterating

  $$\hat{\boldsymbol{x}}_{t+1} = \hat{\boldsymbol{x}}_t + \alpha \cdot \nabla \log p(\hat{\boldsymbol{x}}_t)$$ [stuck in local max]

- At the turn of this decade, the idea of using a good **denoiser** to **synthesize** (*hallucinate*) new images emerged.

- In essence:
  Use the **denoiser** as a **projector** onto the *image manifold*.

- Main (backward) steps:
  1. Start with (or *get to*) a **random noisy** image.
  2. "Climb" to a more probable image by **Langevin Dynamics.**
  3. …and look back at the Bayesian literature of the '60s…surprise!

- Pretend

  We can access the **true/clean** image distribution $p(\boldsymbol{x})$

- Goal

  Draw a sample from $p(\boldsymbol{x})$

Step0: start with a purely random noise image $\widehat{\boldsymbol{x}}_0$

Step1: climb to a more probable image by iterating

$$\widehat{\boldsymbol{x}}_{t+1} = \widehat{\boldsymbol{x}}_t + \alpha \cdot \nabla \log p(\widehat{\boldsymbol{x}}_t) + \beta \cdot \boldsymbol{\eta}_t$$

random perturbation

[Langevin]

- At the turn of this decade, the idea of using a good **denoiser** to **synthesize** (*hallucinate*) new images emerged.

- In essence:
  Use the **denoiser** as a **projector** onto the *image manifold*.

- Main (backward) steps:
  1. Start with (or *get to*) a **random noisy** image.
  2. "Climb" to a more probable image by **Langevin Dynamics.**
  3. …and look back at the Bayesian literature of the '60s…surprise!

- Pretend

  We can access the **true/clean** image distribution $p(\boldsymbol{x})$

- Goal

  Draw a sample from $p(\boldsymbol{x})$

Step0: start with a purely random noise image $\widehat{\boldsymbol{x}}_0$

Step1: climb to a more probable image by iterating

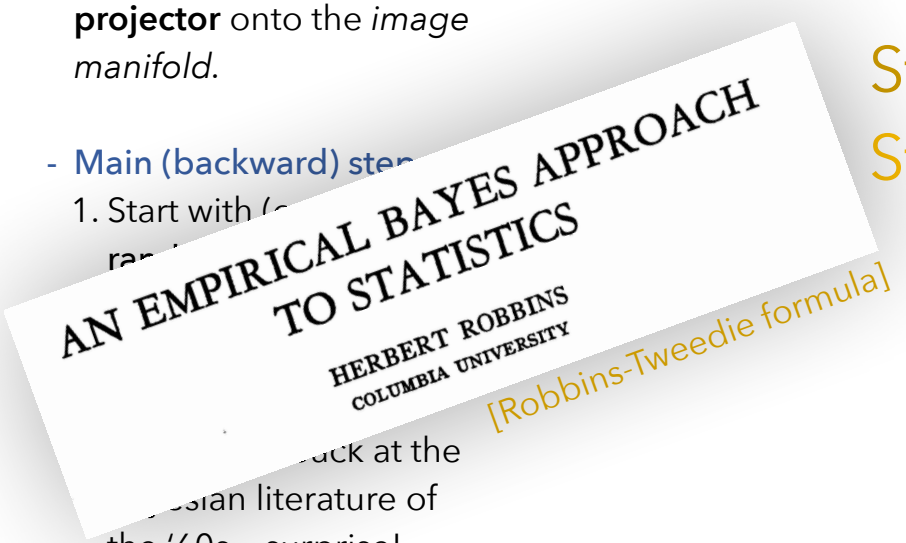
$$\widehat{\boldsymbol{x}}_{t+1} = \widehat{\boldsymbol{x}}_t + \alpha \cdot \nabla \log p(\widehat{\boldsymbol{x}}_t) + \beta \cdot \boldsymbol{\eta}_t \qquad \text{[Langevin]}$$

score function | unknown

- At the turn of this decade, the idea of using a good **denoiser** to **synthesize** (*hallucinate*) new images emerged.

- In essence:
  Use the **denoiser** as a **projector** onto the *image manifold*.

- Main (backward) step
  1. Start with (a
     ran
  2. ack at the
     sian literature of
     the '60s…surprise!


AN EMPIRICAL BAYES APPROACH TO STATISTICS

HERBERT ROBBINS
COLUMBIA UNIVERSITY

[Robbins-Tweedie formula]

- Pretend

  We can access the **true/clean** image distribution $p(\boldsymbol{x})$

- Goal

  Draw a sample from $p(\boldsymbol{x})$

Step0: start with a purely random noise image $\widehat{\boldsymbol{x}}_0$

Step1: climb to a more probable image by iterating

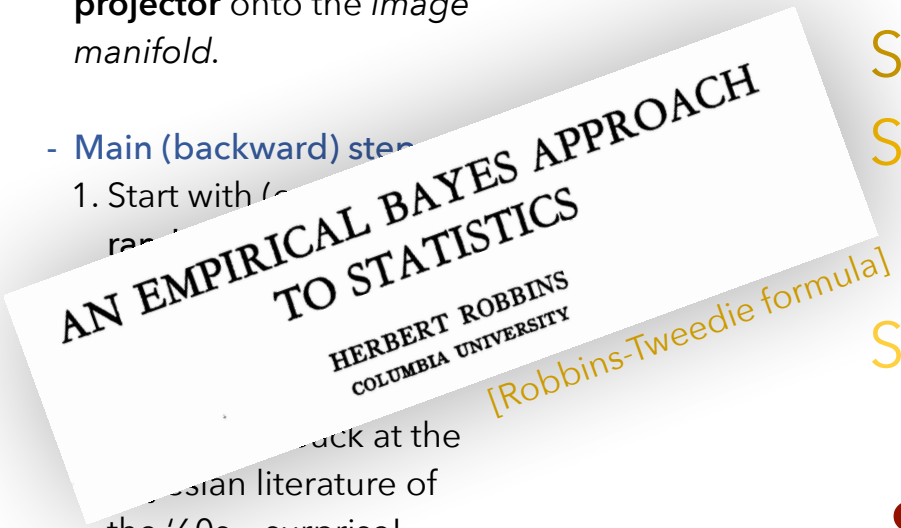
$$\widehat{\boldsymbol{x}}_{t+1} = \widehat{\boldsymbol{x}}_t + \alpha \cdot \nabla \log p(\widehat{\boldsymbol{x}}_t) + \beta \cdot \boldsymbol{\eta}_t \qquad \text{[Langevin]}$$

score function | unknown

- At the turn of this decade, the idea of using a good **denoiser** to **synthesize** (*hallucinate*) new images emerged.

- In essence:
  Use the **denoiser** as a **projector** onto the *image manifold*.

- Main (backward) step
  1. Start with (a
     ran
  2. Look back at the
     Bayesian literature of
     the '60s…surprise!


AN EMPIRICAL BAYES APPROACH TO STATISTICS

HERBERT ROBBINS
COLUMBIA UNIVERSITY

[Robbins-Tweedie formula]

- Pretend
  We can access the **true/clean** image distribution $p(\boldsymbol{x})$
- Goal
  Draw a sample from $p(\boldsymbol{x})$

Step0: start with a purely random noise image $\widehat{\boldsymbol{x}}_0$

Step1: climb to a more probable image by iterating

$$\widehat{\boldsymbol{x}}_{t+1} = \widehat{\boldsymbol{x}}_t + \alpha \cdot \nabla \log p(\widehat{\boldsymbol{x}}_t) + \beta \cdot \boldsymbol{\eta}_t \qquad \text{[Langevin]}$$

score function | unknown

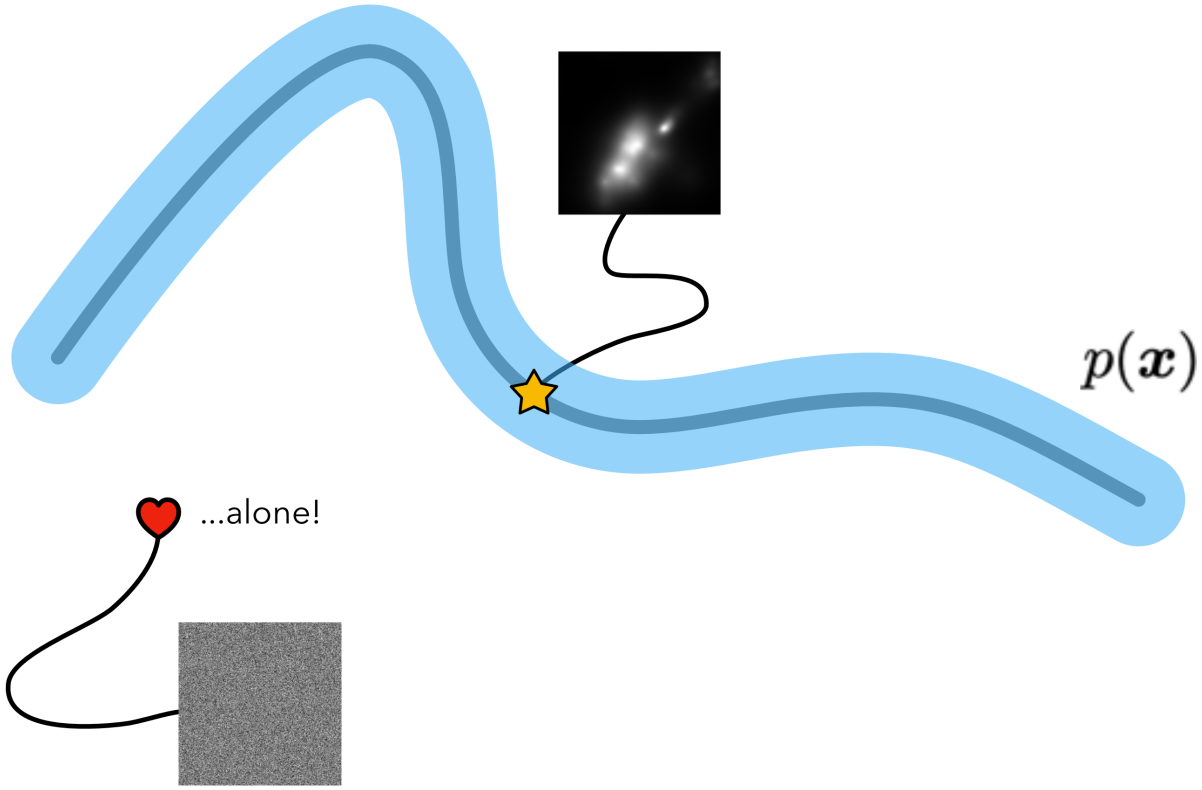
Step2: look at the Empirical Bayes literature to see that

$$\nabla \log p(\widehat{\boldsymbol{x}}_t) \propto \left[\widehat{\boldsymbol{x}}_t - \mathrm{D}(\widehat{\boldsymbol{x}}_t, \sigma)\right] \qquad \text{[small } \sigma\text{]}$$

implicit link between MMSE **denoisers** and **priors**

- "Technical" caveat
  In practice, instead of the plain Langevin with a fixed (and small) noise level, we use the Annealed Langevin that considers a sequence of blurred priors associated with a decreasing sequence of noise levels:

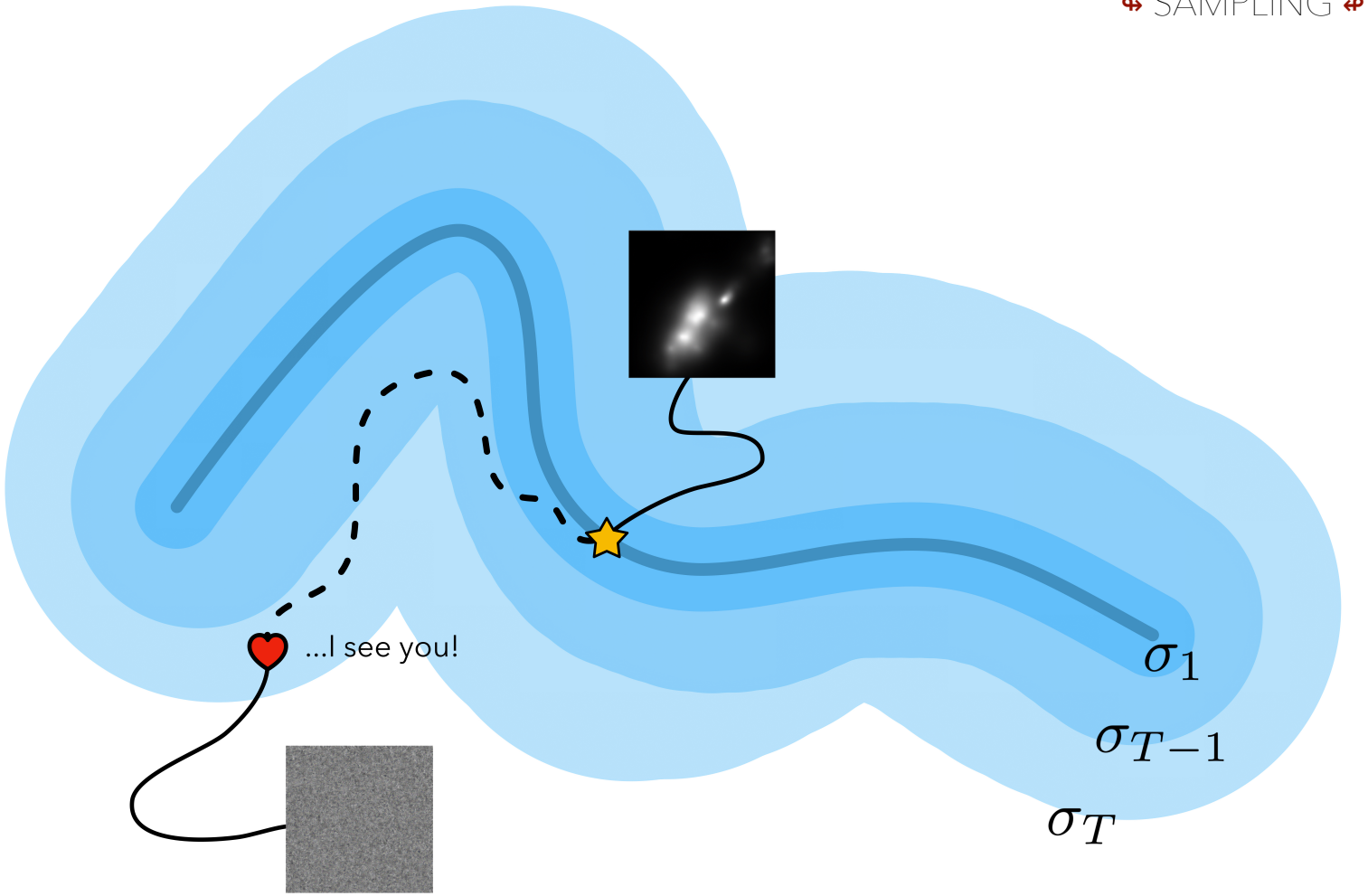
$$\sigma_T > \cdots > \sigma_1$$



- "Technical" caveat

  In practice, instead of the plain Langevin with a fixed (and small) noise level, we use the Annealed Langevin that considers a sequence of blurred priors associated with a decreasing sequence of noise levels:

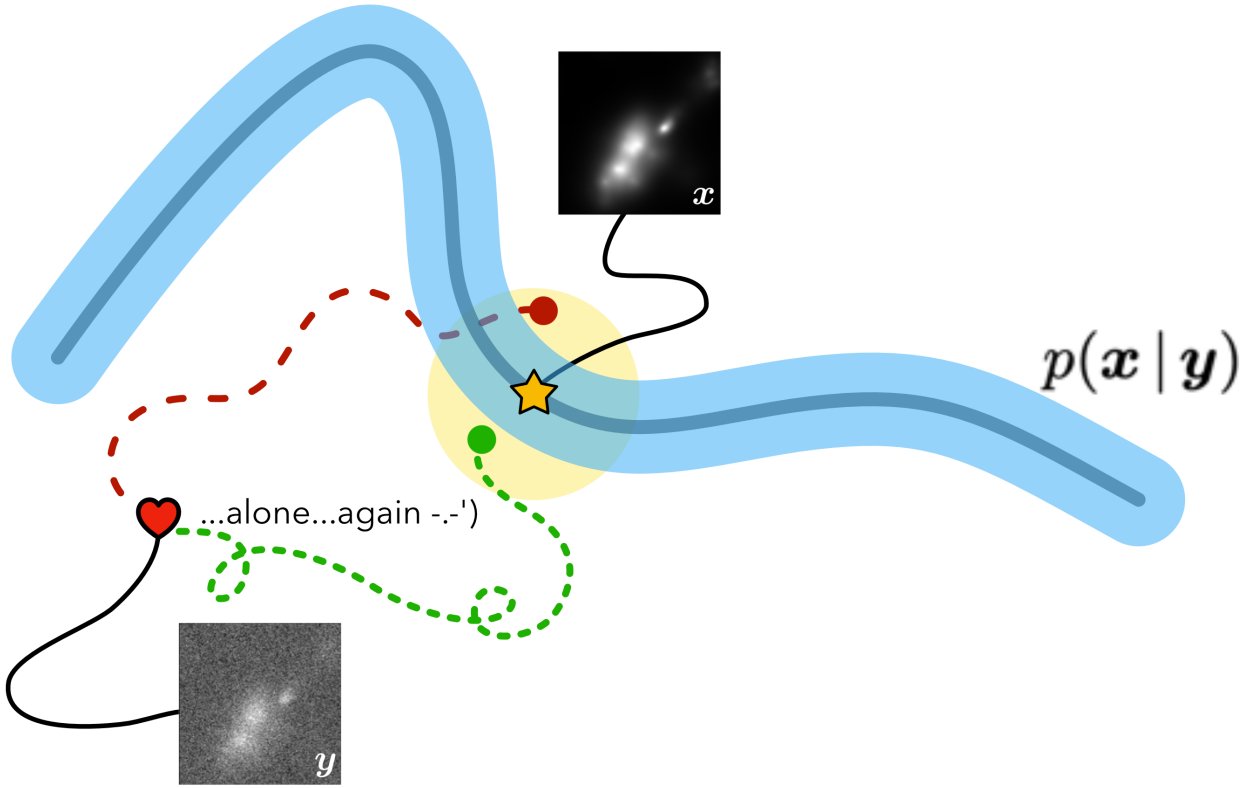
$$\sigma_T > \cdots > \sigma_1$$



- Denoising by Sampling:
  The Annealed Langevin algorithm can be twisted to design a **Stochastic Image Denoiser** that recover a sequence of gradually less noisy images by targeting the posterior distribution

$$p(\boldsymbol{x} \mid \boldsymbol{y})$$

  Solution is **not** unique.



- ...and beyond!

  Pretty much the same Langevin-based paradigm can be adapted to deal with a variety of **inverse problems** in order to recover a signal from **corrupted measurements**:

$$\boldsymbol{y} = \mathsf{H}\boldsymbol{x} + \boldsymbol{\epsilon}$$

  - De-blurring
  - In-Painting
  - De-Mosaicing
  - Super-Resolution

- ...and beyond!

  Pretty much the same Langevin-based paradigm can be adapted to deal with a variety of **inverse problems** in order to recover a signal from **corrupted measurements**:

$$\boldsymbol{y} = \mathsf{H}\boldsymbol{x} + \boldsymbol{\epsilon}$$

  - De-blurring
  - In-Painting
  - De-Mosaicing
  - Super-Resolution

- Power is nothing without control



PNRR Activity: Topological Data Analysis for Astronomical Imaging

MACHINE LEARNING FOR ASTROPHYSICS
3RD EDITION



TUESDAY 01/09

THURSDAY 03/09

FRIDAY 04/09
(GenAI Day)

MACHINE LEARNING FOR ASTROPHYSICS
3RD EDITION



Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
INAF ISTITUTO NAZIONALE DI ASTROFISICA

TUESDAY 01/09

Edoardo Giancarli | 11:10 - 11:30
IROS Diffusion Imaging for the LEM-X Observatory

Anja Schmit | 14:30 - 14:50
Using Flow Matching to study the sub-burst structure of short astronomical transients

THURSDAY 03/09

Anumanchi Agastya Likhit | 12:15 - 12:20
Score-Based Diffusion for Low-ℓ Primordial CMB B-Mode Reconstruction

Tirthankar De | 12:20 - 12:25
Robust Cosmological Parameter Inference from Galaxy Properties using Conditional Normalizing Flows

Boris Khanikati | 15:00 - 15:20
Anomalous Dark Matter Halos with Normalizing Flows

FRIDAY 04/09
(GenAI Day)

Mehdi Noor | 09:00 - 09:20
How do generative diffusion models learn cosmic web environments?

Luca Fontana | 14:00 - 14:20
Deep Generative Priors for Interferometry: Non-Parametric Forward Modelling of the Sunyaev-Zeldovich Effect

MACHINE LEARNING FOR ASTROPHYSICS
3RD EDITION



Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
INAF ISTITUTO NAZIONALE DI ASTROFISICA

TUESDAY 01/09

Edoardo Giancarli | 11:10 - 11:30
IROS Diffusion Imaging for the LEM-X Observatory

Anja Schmit | 14:30 - 14:50
Using Flow Matching to study the sub-burst structure of short astronomical transients

THURSDAY 03/09

Anumanchi Agastya Likhit | 12:15 - 12:20
Score-Based Diffusion for Low-ℓ Primordial CMB B-Mode Reconstruction

Tirthankar De | 12:20 - 12:25
Robust Cosmological Parameter Inference from Galaxy Properties using Conditional Normalizing Flows

Boris Khanikati | 15:00 - 15:20
Anomalous Dark Matter Halos with Normalizing Flows

FRIDAY 04/09
(GenAI Day)

Mehdi Noor | 09:00 - 09:20
How do generative diffusion models learn cosmic web environments?

Luca Fontana | 14:00 - 14:20
Deep Generative Priors for Interferometry: Non-Parametric Forward Modelling of the Sunyaev-Zeldovich Effect

MACHINE LEARNING FOR ASTROPHYSICS
3RD EDITION

↬ 2nd pass ↫

From **denoising** to **sampling** and back...







Flavor 1: Diffusion, DDPM	$x_t = \sqrt{\overline{\alpha}_t} x_0 + \sqrt{1 - \overline{\alpha}_t}\epsilon$ **Forward** Process $x_{t-1} = \frac{1}{\sqrt{\alpha_t}} \left(x_t - \frac{1 - \alpha_t}{\sqrt{1 - \overline{\alpha}_t}} \epsilon_\theta(x_t, t) \right) + \sigma_t \epsilon$

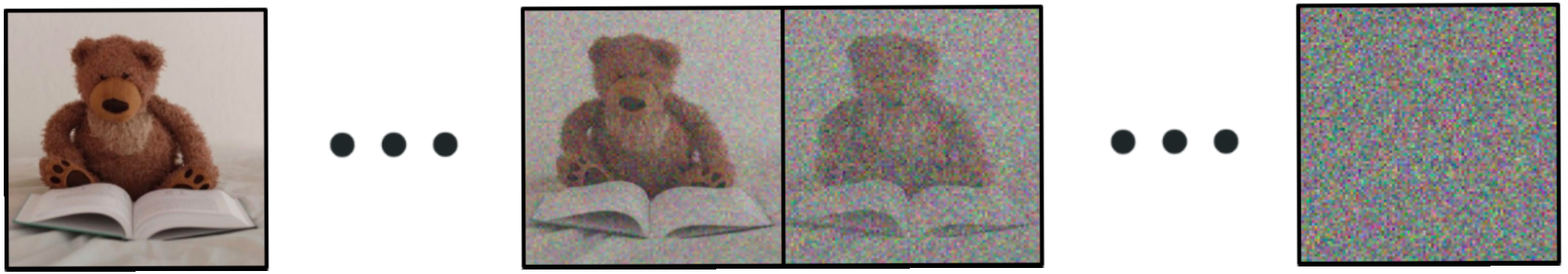


Flavor 1: Diffusion, DDPM	$x_t = \sqrt{\overline{\alpha}_t} x_0 + \sqrt{1-\overline{\alpha}_t}\epsilon$	**Forward** Process
	$x_{t-1} = \frac{1}{\sqrt{\alpha_t}}\left(x_t - \frac{1-\alpha_t}{\sqrt{1-\overline{\alpha}_t}}\epsilon_\theta(x_t, t)\right) + \sigma_t \epsilon$	**Backward** or **Generation** Process



Flavor 1: Diffusion, DDPM	$x_t = \sqrt{\overline{\alpha}_t} x_0 + \sqrt{1-\overline{\alpha}_t}\epsilon$ **Forward** Process $x_{t-1} = \frac{1}{\sqrt{\alpha_t}}\left(x_t - \frac{1-\alpha_t}{\sqrt{1-\overline{\alpha}_t}}\epsilon_\theta(x_t, t)\right) + \sigma_t\epsilon$ **Backward** or **Generation** Process

FORWARD



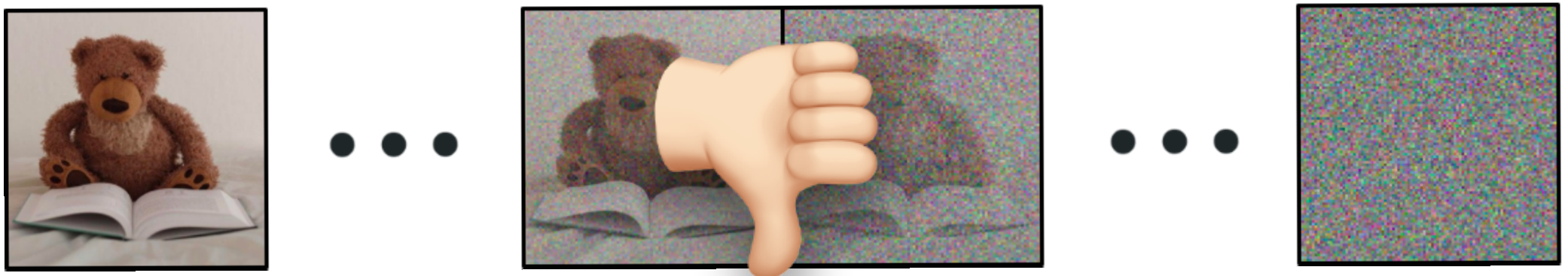
BACKWARD

SOFT TEDDY BEAR?



Flavor 1: Diffusion, DDPM	$x_t = \sqrt{\overline{\alpha}_t} x_0 + \sqrt{1-\overline{\alpha}_t}\epsilon$ **Forward** Process $x_{t-1} = \frac{1}{\sqrt{\alpha_t}}\left(x_t - \frac{1-\alpha_t}{\sqrt{1-\overline{\alpha}_t}}\epsilon_\theta(x_t, t)\right) + \sigma_t\epsilon$ **Backward** or **Generation** Process

FORWARD



BACKWARD

SOFT TEDDY BEAR?



Flavor 1:
Diffusion, DDPM

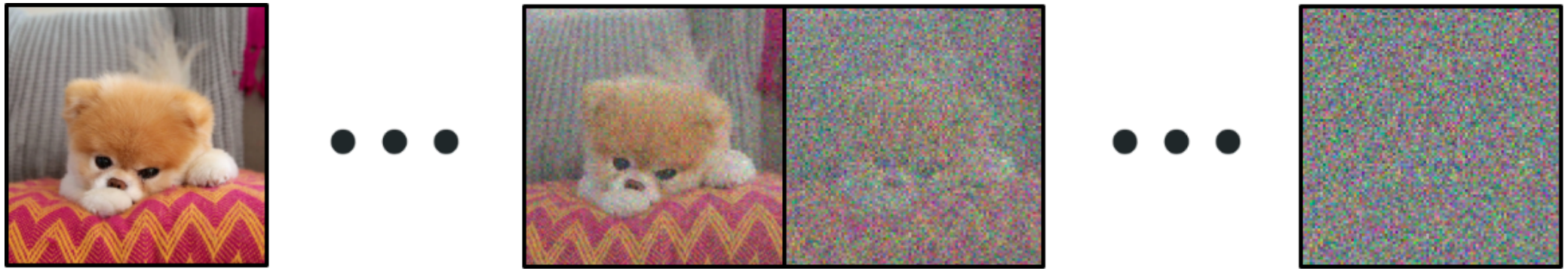
$$x_t = \sqrt{\overline{\alpha}_t} x_0 + \sqrt{1 - \overline{\alpha}_t}\epsilon$$

Forward Process

$$x_{t-1} = \frac{1}{\sqrt{\alpha_t}} \left(x_t - \frac{1 - \alpha_t}{\sqrt{1 - \overline{\alpha}_t}} \epsilon_\theta(x_t, t) \right) + \sigma_t \epsilon$$

Backward or **Generation** Process

FORWARD



BACKWARD

CUTE PUPPY?



Flavor 1:
Diffusion, DDPM

$$x_t = \sqrt{\overline{\alpha}_t} x_0 + \sqrt{1 - \overline{\alpha}_t}\epsilon$$

Forward Process

$$x_{t-1} = \frac{1}{\sqrt{\alpha_t}} \left(x_t - \frac{1 - \alpha_t}{\sqrt{1 - \overline{\alpha}_t}} \epsilon_\theta(x_t, t) \right) + \sigma_t \epsilon$$

Backward or **Generation** Process

FORWARD



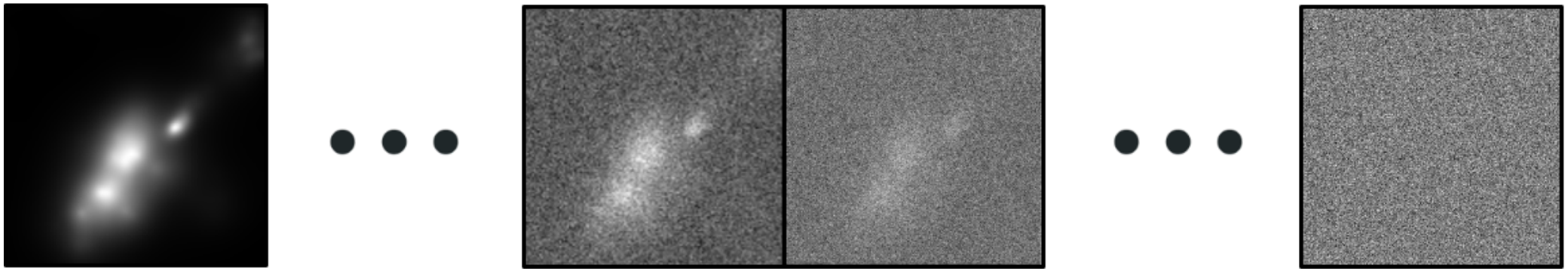
BACKWARD

CUTE PUPPY?



Flavor 1: Diffusion, DDPM	$x_t = \sqrt{\overline{\alpha}_t} x_0 + \sqrt{1 - \overline{\alpha}_t}\epsilon$ **Forward** Process $x_{t-1} = \frac{1}{\sqrt{\alpha_t}} \left(x_t - \frac{1 - \alpha_t}{\sqrt{1 - \overline{\alpha}_t}} \epsilon_\theta(x_t, t) \right) + \sigma_t \epsilon$ **Backward** or **Generation** Process

FORWARD



BACKWARD

GRAYSCALE GALAXY!



Flavor 1:
Diffusion, DDPM

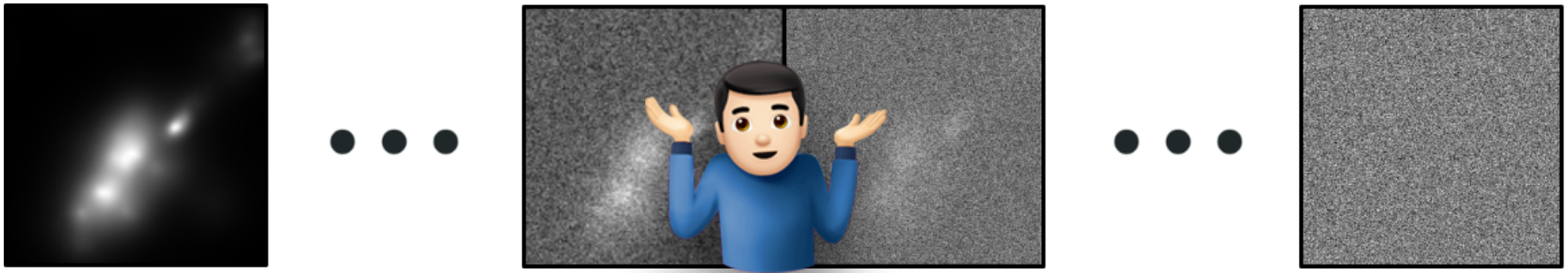
$$x_t = \sqrt{\overline{\alpha}_t} x_0 + \sqrt{1 - \overline{\alpha}_t}\epsilon$$

Forward Process

$$x_{t-1} = \frac{1}{\sqrt{\alpha_t}} \left(x_t - \frac{1 - \alpha_t}{\sqrt{1 - \overline{\alpha}_t}} \epsilon_\theta(x_t, t) \right) + \sigma_t \epsilon$$

Backward or **Generation** Process

FORWARD



BACKWARD

GRAYSCALE GALAXY!



Flavor 1:
Diffusion, DDPM

$$x_t = \sqrt{\overline{\alpha}_t} x_0 + \sqrt{1 - \overline{\alpha}_t}\epsilon$$

$$x_{t-1} = \frac{1}{\sqrt{\alpha_t}}\left(x_t - \frac{1-\alpha_t}{\sqrt{1-\overline{\alpha}_t}}\epsilon_\theta(x_t, t)\right) + \sigma_t \epsilon$$

$\epsilon_\theta(x_t, t)$ (error)



Flavor 1:
Diffusion, DDPM

$$x_t = \sqrt{\overline{\alpha}_t} x_0 + \sqrt{1 - \overline{\alpha}_t}\epsilon$$

$$x_{t-1} = \frac{1}{\sqrt{\alpha_t}}\left(x_t - \frac{1-\alpha_t}{\sqrt{1-\overline{\alpha}_t}}\epsilon_\theta(x_t, t)\right) + \sigma_t \epsilon$$

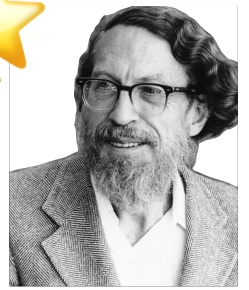
$\epsilon_\theta(x_t, t)$
(error)

Flavor 2:
Score Matching, SDE

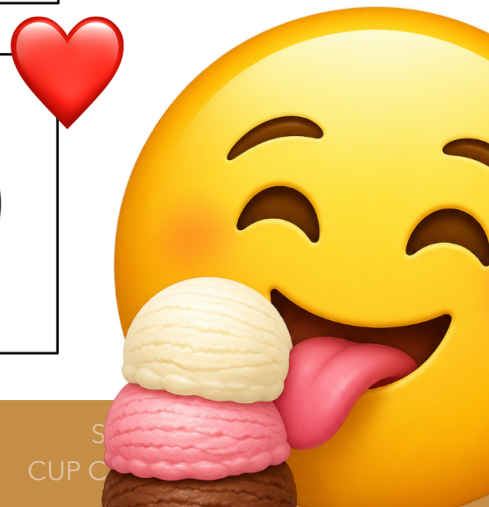
$$dx = f(x_t, t)dt + g(t)dW$$

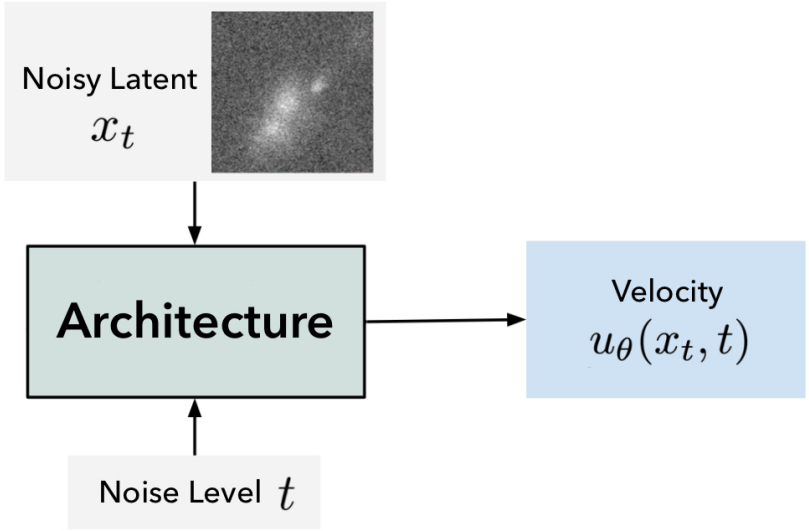
$$dx = \left[f(x_t, t) - g(t)^2 s_\theta(x_t, t)\right]dt + g(t)d\overline{W}$$

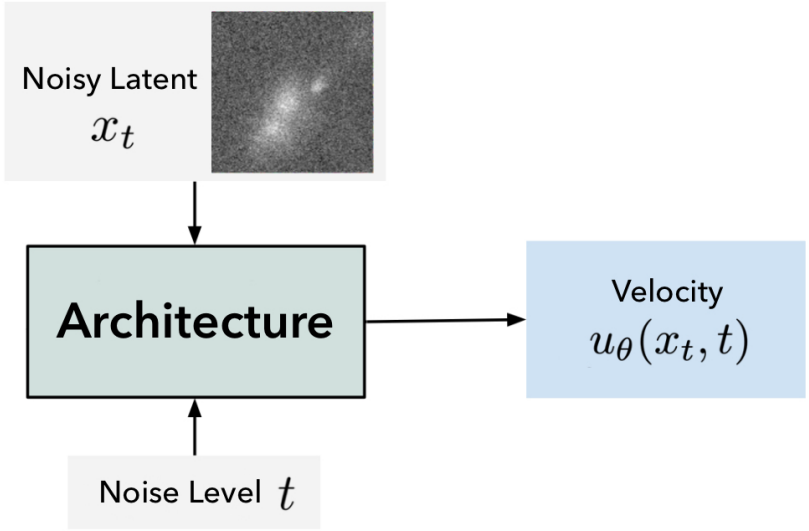
$s_\theta(x_t, t)$
(Stein' score)

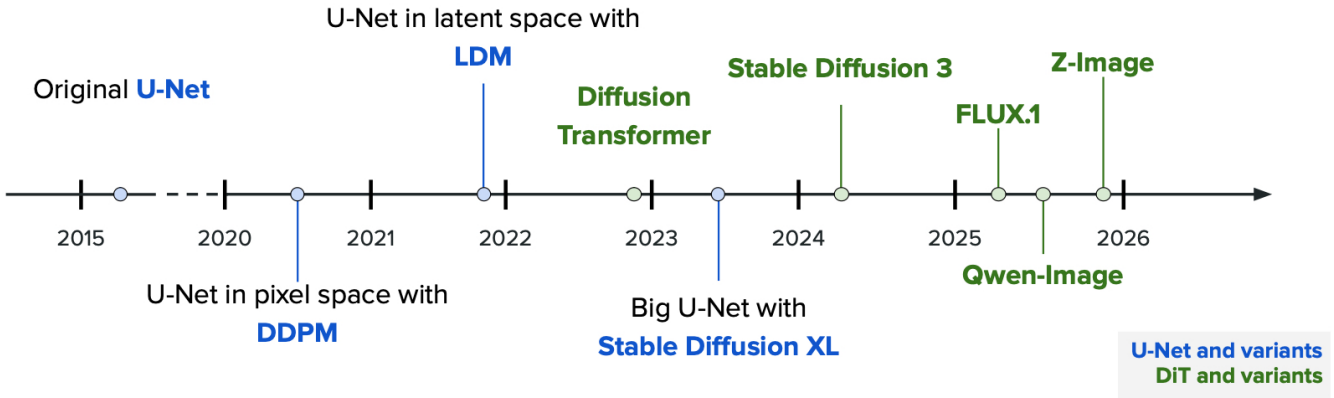

(Charles)

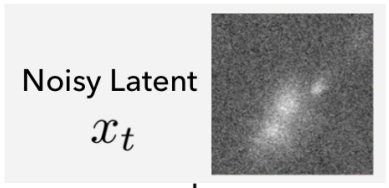
Flavor 1: Diffusion, DDPM	$x_t = \sqrt{\overline{\alpha}_t} x_0 + \sqrt{1-\overline{\alpha}_t}\epsilon$ $x_{t-1} = \frac{1}{\sqrt{\alpha_t}}\left(x_t - \frac{1-\alpha_t}{\sqrt{1-\overline{\alpha}_t}}\epsilon_\theta(x_t,t)\right) + \sigma_t\epsilon$	$\epsilon_\theta(x_t,t)$ (error)
Flavor 2: Score Matching, SDE	$dx = f(x_t,t)dt + g(t)dW$ $dx = \left[f(x_t,t) - g(t)^2 s_\theta(x_t,t)\right]dt + g(t)d\overline{W}$	$s_\theta(x_t,t)$ (Stein' score)
Flavor 3: Flow Matching	$x_t = (1-t)x_0 + tx_1$ $x_{t+\Delta t} = x_t + u_\theta(x_t,t)\Delta t$	$u_\theta(x_t,t)$ (velocity)

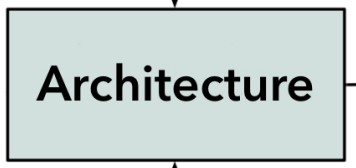



Noisy Latent
x_t
Architecture
Velocity
$u_\theta(x_t, t)$
Noise Level t

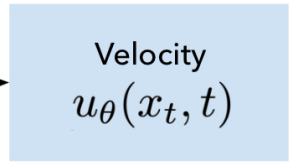

Noisy Latent
x_t
Architecture
Velocity
$u_\theta(x_t, t)$
Noise Level t

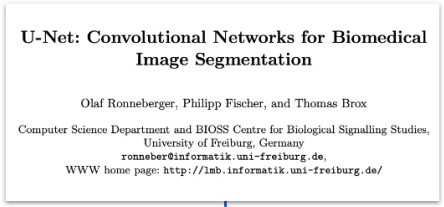

Original U-Net
U-Net in latent space with LDM
U-Net in pixel space with DDPM
Diffusion Transformer
Big U-Net with Stable Diffusion XL
Stable Diffusion 3
FLUX.1
Z-Image
Qwen-Image
2015
2020
2021
2022
2023
2024
2025
2026
U-Net and variants
DiT and variants





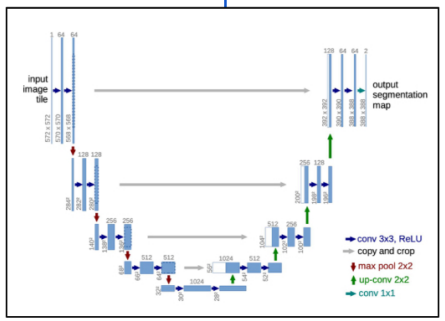


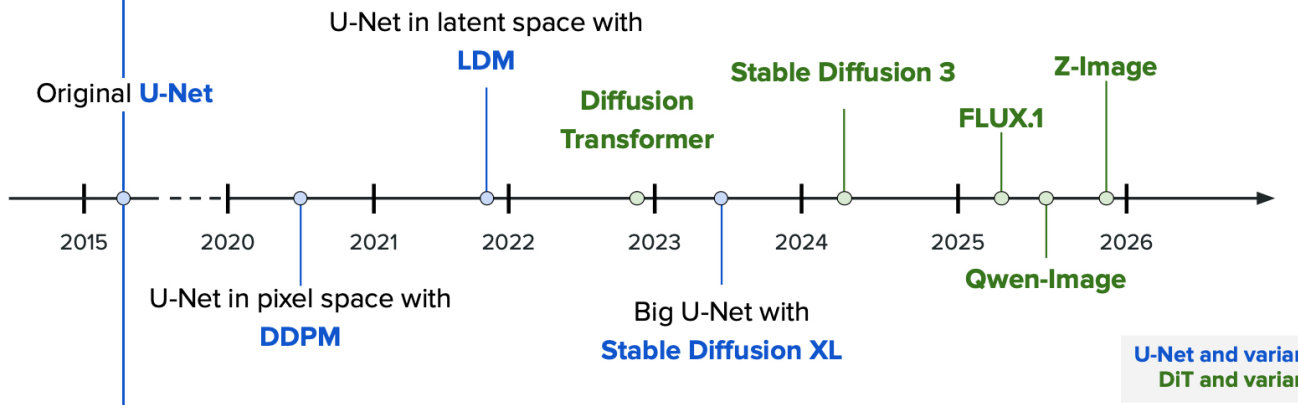


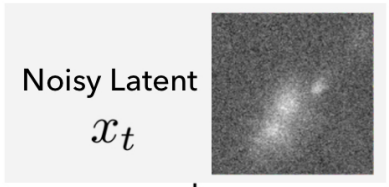

U-Net: Convolutional Networks for Biomedical Image Segmentation

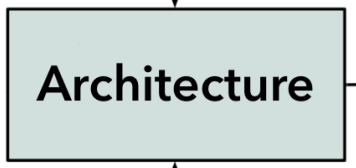
Olaf Ronneberger, Philipp Fischer, and Thomas Brox

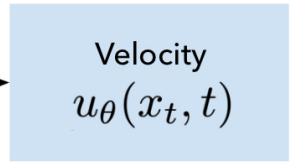
Computer Science Department and BIOSS Centre for Biological Signalling Studies, University of Freiburg, Germany
ronneber@informatik.uni-freiburg.de,
WWW home page: http://lmb.informatik.uni-freiburg.de/



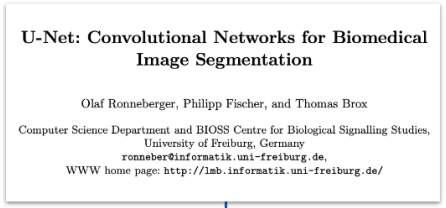



Noisy Latent
x_t

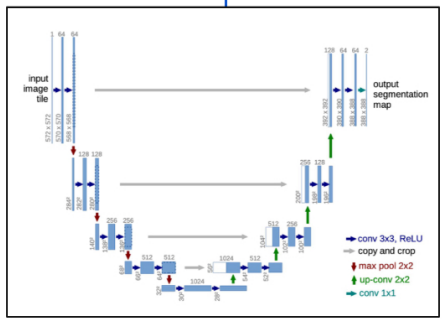

Architecture

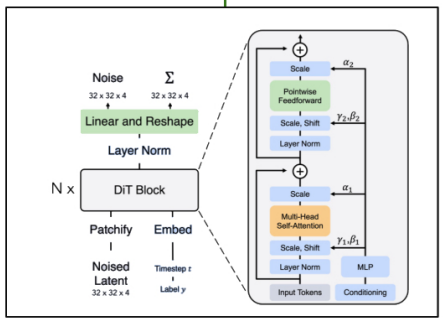

Velocity
$u_\theta(x_t, t)$

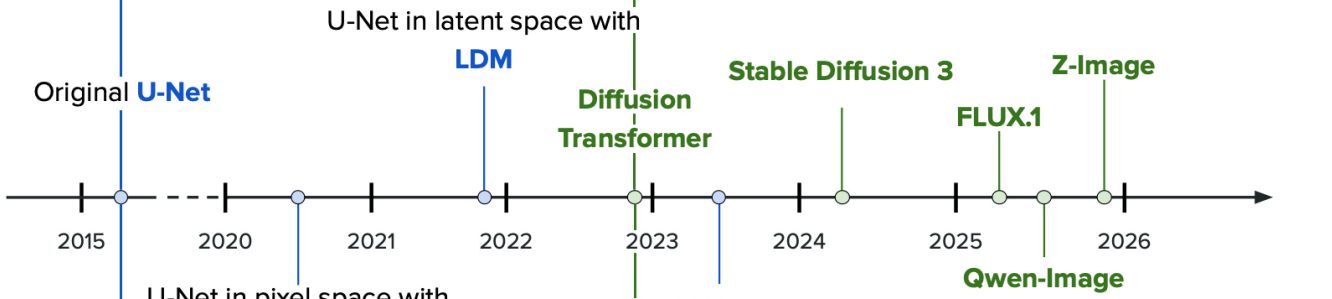

Noise Level t


U-Net: Convolutional Networks for Biomedical Image Segmentation
Olaf Ronneberger, Philipp Fischer, and Thomas Brox
Computer Science Department and BIOSS Centre for Biological Signalling Studies, University of Freiburg, Germany
ronneber@informatik.uni-freiburg.de,
WWW home page: http://lmb.informatik.uni-freiburg.de/


Scalable Diffusion Models with Transformers
William Peebles*
UC Berkeley
Saining Xie
New York University


input image tile
output segmentation map
conv 3x3, ReLU
copy and crop
max pool 2x2
up-conv 2x2
conv 1x1


Noise
Σ
Linear and Reshape
Layer Norm
N x
DiT Block
Patchify
Embed
Noised Latent
Scale
Pointwise Feedforward
Scale, Shift
Layer Norm
Multi-Head Self-Attention
MLP
Input Tokens
Conditioning


Original U-Net
U-Net in pixel space with DDPM
U-Net in latent space with LDM
Diffusion Transformer
Big U-Net with Stable Diffusion XL
Stable Diffusion 3
FLUX.1
Z-Image
Qwen-Image
2015
2020
2021
2022
2023
2024
2025
2026
U-Net and variants
DiT and variants

TIMELINE & UNIFICATION

Flavor 1:
Diffusion, DDPM

Flavor 2:
Score Matching, SDE

TIMELINE & UNIFICATION

Flavor 1: Diffusion, DDPM

Flavor 2: Score Matching, SDE

Intuition

Deep Unsupervised Learning using Nonequilibrium Thermodynamics

Jascha Sohl-Dickstein — Stanford University — JASCHA@STANFORD.EDU
Eric A. Weiss — University of California, Berkeley — EAWEISS@BERKELEY.EDU

2015

$\mathcal{L}_{\text{DDPM}}$

Denoising Diffusion Probabilistic Models

Jonathan Ho — UC Berkeley — jonathanho@berkeley.edu
Ajay Jain — UC Berkeley — ajayj@berkeley.edu
Pieter Abbeel — UC Berkeley — pabbeel@cs.berkeley.edu

2020

SDE

SCORE-BASED GENERATIVE MODELING THROUGH STOCHASTIC DIFFERENTIAL EQUATIONS

Yang Song* — Stanford University — yangsong@cs.stanford.edu
Jascha Sohl-Dickstein — Google Brain — jaschasd@google.com
Diederik P. Kingma — Google Brain — durk@google.com
Abhishek Kumar — Google Brain — abhishk@google.com
Stefano Ermon — Stanford University — ermon@cs.stanford.edu
Ben Poole — Google Brain — pooleb@google.com

2020

2006

2010

2019

Principle

Estimation of Non-Normalized Statistical Models by Score Matching

Aapo Hyvärinen — *Helsinki Institute for Information Technology (BRU)*, *Department of Computer Science*, *FIN-00014 University of Helsinki, Finland* — AAPO.HYVARINEN@HELSINKI.FI

$\mathcal{L}_{\text{DSM}}$

A Connection Between Score Matching and Denoising Autoencoders

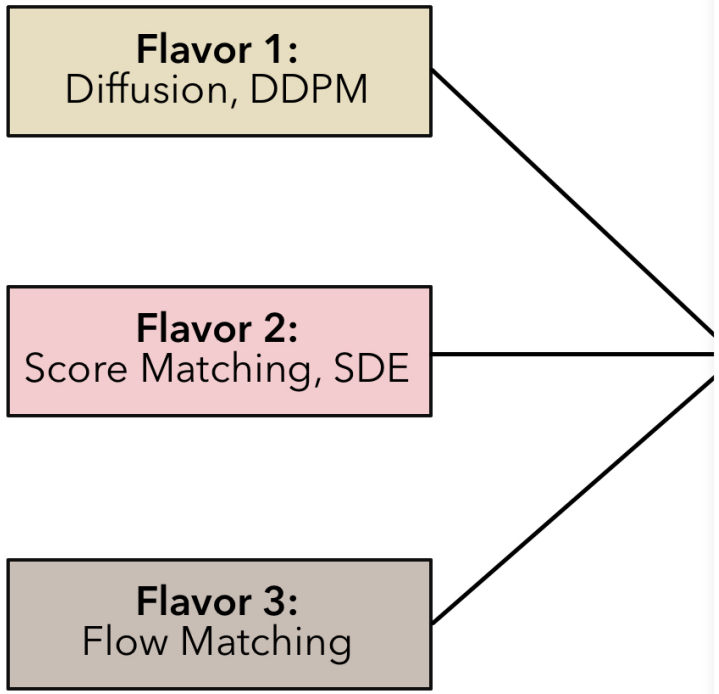
Pascal Vincent — *vincentp@iro.umontreal.ca* — *Dept. IRO, Université de Montréal,*

$\mathcal{L}_{\text{NCSN}}$

Generative Modeling by Estimating Gradients of the Data Distribution

Yang Song — Stanford University — yangsong@cs.stanford.edu
Stefano Ermon — Stanford University — ermon@cs.stanford.edu

TIMELINE & UNIFICATION



Journal of Machine Learning Research 26 (2025) 1-80 Submitted 12/23; Revised 9/25; Published 9/25

Stochastic Interpolants: A Unifying Framework for Flows and Diffusions

Michael S. Albergo* ALBERGO@NYU.EDU
Center for Cosmology and Particle Physics
New York University
New York, NY 10012, USA

Nicholas M. Boffi* BOFFI@CIMS.NYU.EDU
Courant Institute of Mathematical Sciences
New York University
New York, NY 10012, USA

Eric Vanden-Eijnden EVE2@CIMS.NYU.EDU
Courant Institute of Mathematical Sciences
New York University
New York, NY 10012, USA

Editor: Maxim Raginsky

OPEN WEIGHTS LEADERBOARD

Artificial Analysis Text to Image Leaderboard Artificial Analysis

	Range	Creator	Model	Elo	95% CI	Samples	Released	API Pricing [1]
1	1	HiDream	HiDream-O1-Image-Dev-2604 Open Weights	1,187	−9/9	3,834	May 2026	No API available
2	2–4	Alibaba	Qwen Image Max 2512 Open Weights	1,159	−9/9	3,992	Dec 2025	$20.0 /1k imgs
3	2–4	Black Forest Labs	FLUX.2 [dev] Open Weights	1,159	−9/9	3,761	Nov 2025	$12.0 /1k imgs
4	2–4	Fal	FLUX.2 [dev] Turbo Open Weights	1,159	−9/9	4,378	Dec 2025	$8.0 /1k imgs
5	5	Fal	FLUX.2 [dev] Flash Open Weights	1,143	−9/9	4,361	Dec 2025	$5.0 /1k imgs

Open-weights models | May'26

OPEN WEIGHTS LEADERBOARD

Artificial Analysis Text to Image Leaderboard Artificial Analysis

	Range	Creator	Model	Elo	95% CI			
3	2–4	Black Forest Labs	FLUX.2 [dev] Open Weights	1,159	−9/9	3,761	Nov 2025	$12.0 /1k imgs
4	2–4	Fal	FLUX.2 [dev] Turbo Open Weights	1,159	−9/9	4,378	Dec 2025	$8.0 /1k imgs
5	5	Fal	FLUX.2 [dev] Flash Open Weights	1,143	−9/9	4,361	Dec 2025	$5.0 /1k imgs

Black Forest Labs

NOVEMBER 25, 2025

FLUX.2: Analyzing and Enhancing the Latent Space of FLUX – Representation Comparison

- **Flavor**: Rectified flow
- **Architecture**: Hybrid MM-DiT
- **Latent space**: VAE
- **Embeddings**: Mistral 3

OPEN WEIGHTS LEADERBOARD

Artificial Analysis Text to Image Leaderboard Artificial Analysis

	Range	Creator	Model	Elo	95% CI			
2	2–4	Alibaba	Qwen Image Max 2512 Open Weights	1,159	−9/9	3,992	Dec 2025	$20.0 /1k imgs

Qwen August 5, 2025

Qwen-Image Technical Report

Qwen Team

https://qwen.ai
https://huggingface.co/Qwen/Qwen-Image
https://modelscope.cn/models/Qwen/Qwen-Image
https://github.com/QwenLM/Qwen-Image

- **Flavor**: Flow matching
- **Architecture**: MM-DiT
- **Latent space**: VAE
- **Embeddings**: Qwen2.5VL

OPEN WEIGHTS LEADERBOARD

■ Artificial Analysis Text to Image Leaderboard Artificial Analysis

	Range	Creator	Model	Elo	95% CI	Samples	Released	API Pricing [1]
1	1	HiDream	HiDream-O1-Image-Dev-2604 Open Weights	1,187	−9/9	3,834	May 2026	No API available



- **Flavor**: Flow matching
- **Architecture**: Transformer
- **Latent space**: No VAE
- **Embeddings**: No pre-trained encoder

Now...

Finanziato dall'Unione europea
NextGenerationEU

Ministero dell'Università e della Ricerca

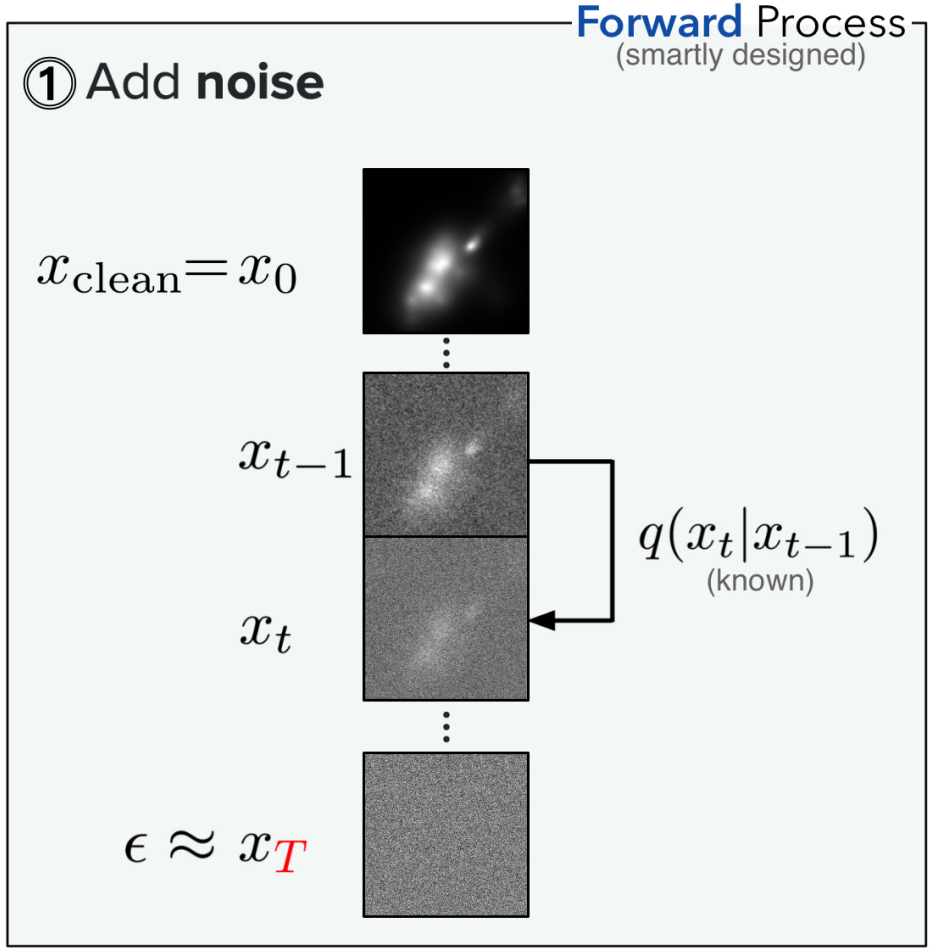
Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA

INAF
ISTITUTO NAZIONALE DI ASTROFISICA

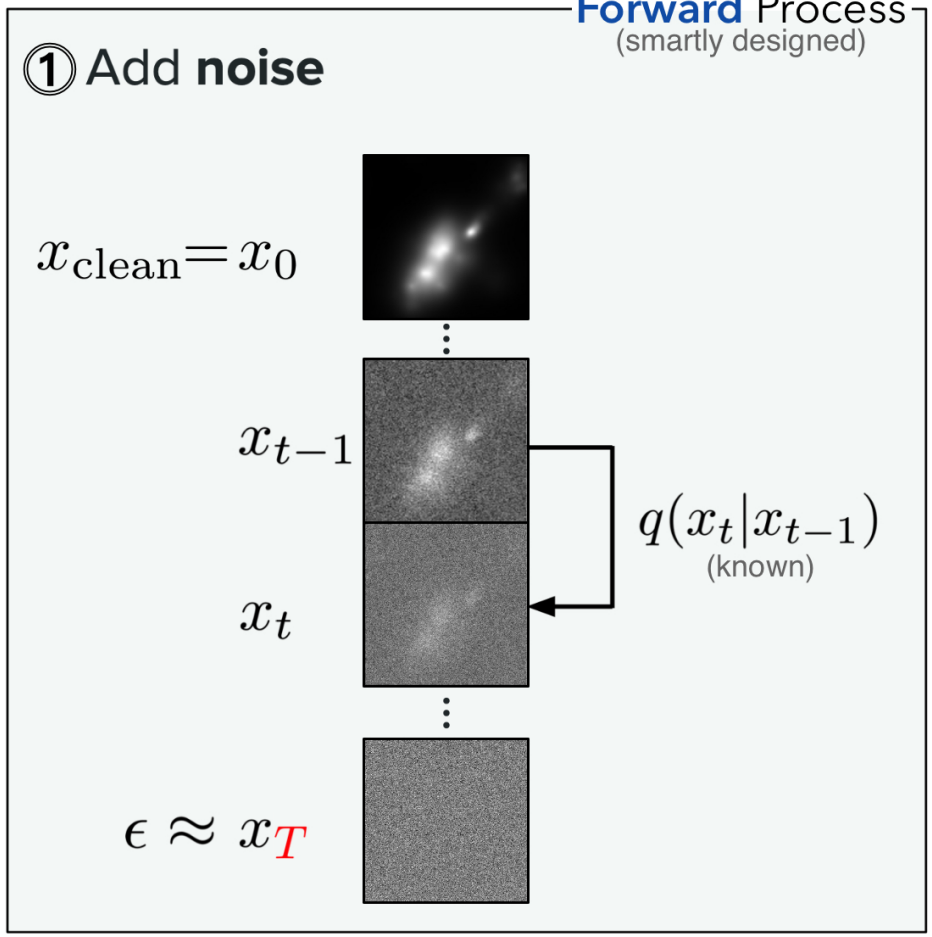
Source: xkcd | https://www.youtube.com/watch?v=r7B0MzvwDcA



Flavor 1:
Diffusion, DDPM



Flavor 1:
Diffusion, DDPM

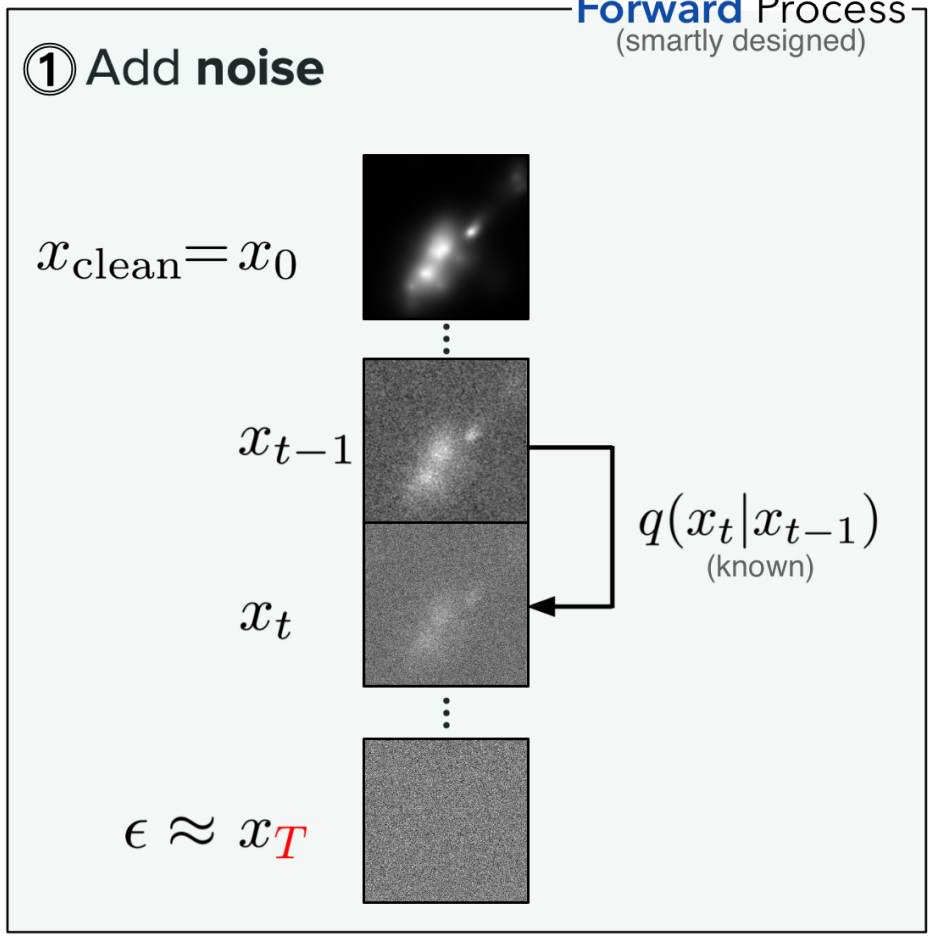


$$q(x_t|x_{t-1}) = \mathrm{N}(\sqrt{1-\beta_t}x_{t-1}, \beta_t\mathbb{I})$$

with $0 \leqslant \beta_1 < \beta_2 < \ldots < \beta_T \leqslant 1$

(noise schedule)

Flavor 1:
Diffusion, DDPM



$$q(x_t|x_{t-1}) = \mathrm{N}(\sqrt{1-\beta_t}x_{t-1}, \beta_t\mathbb{I})$$

with $0 \leqslant \beta_1 < \beta_2 < \ldots < \beta_T \leqslant 1$

(noise schedule)

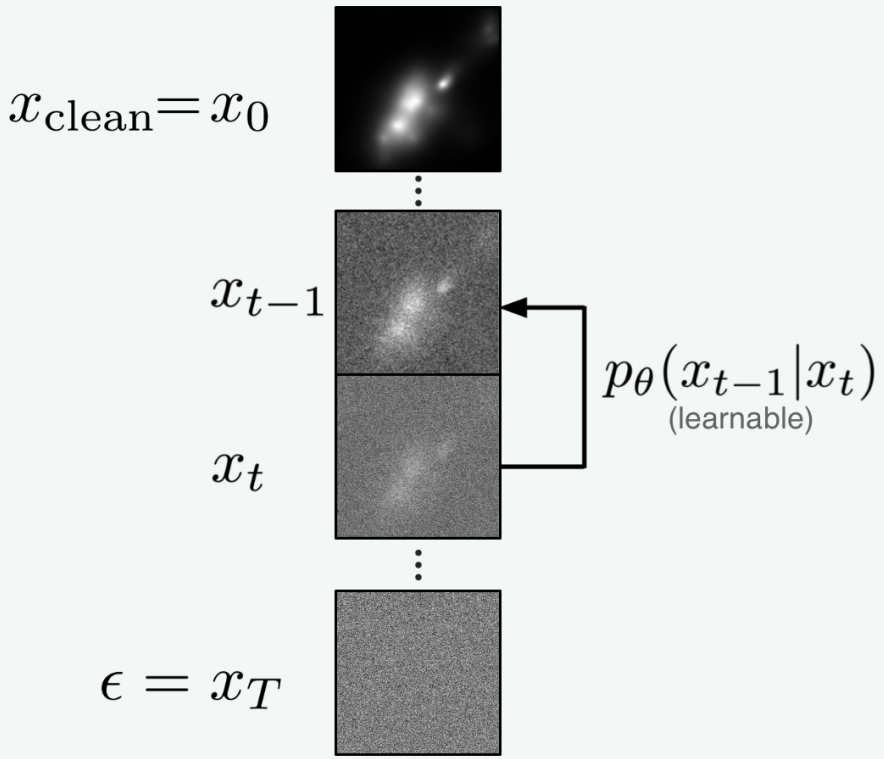
$$\boxed{q(x_t|x_0) = \mathrm{N}(\sqrt{\overline{\alpha}_t}x_0, (1-\overline{\alpha}_t)\mathbb{I})}$$

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

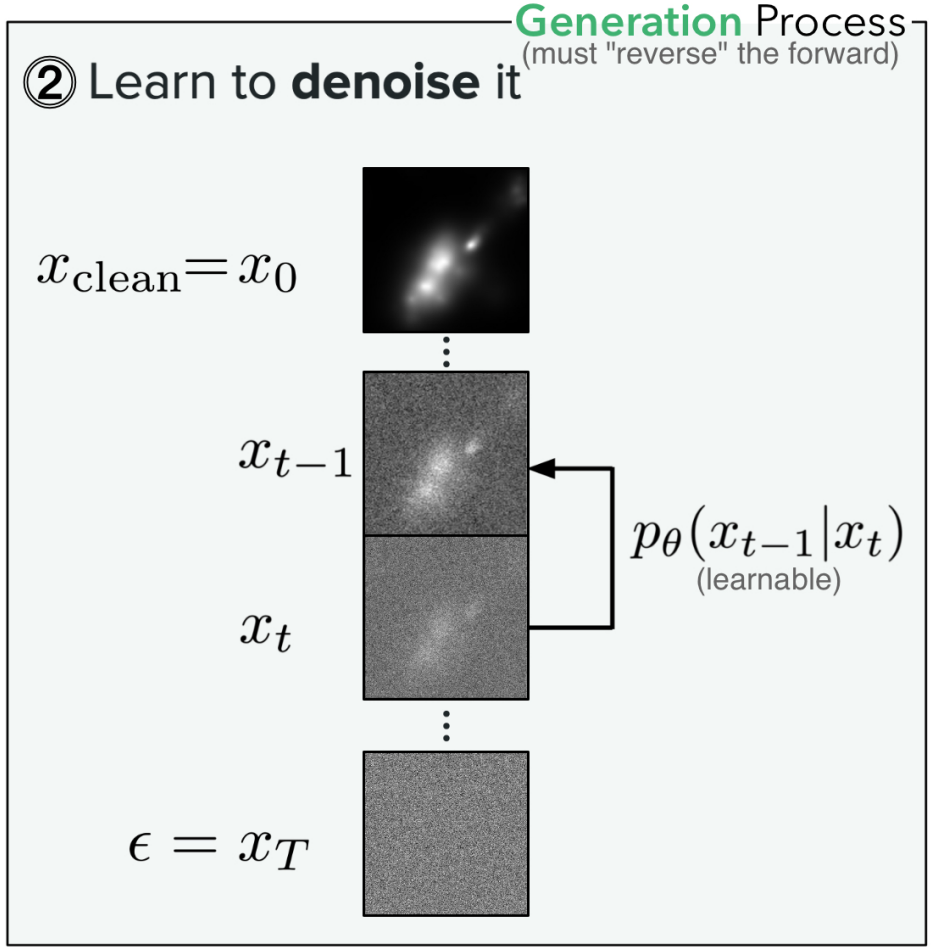
Flavor 1:
Diffusion, DDPM

Generation Process
(must "reverse" the forward)

② Learn to **denoise** it



Flavor 1:
Diffusion, DDPM



STRATEGY

A. Get (variational) **lower bound** on log-likelihood
   - Jensen inequality ↭ E(vidence)L(ower)BO(und)

B. Expand terms of lower bound
   - $\mathrm{KL}\left(q(x_{t-1}|x_t, x_0) \,\middle\|\, p_\theta(x_{t-1}|x_t)\right)$ pops up

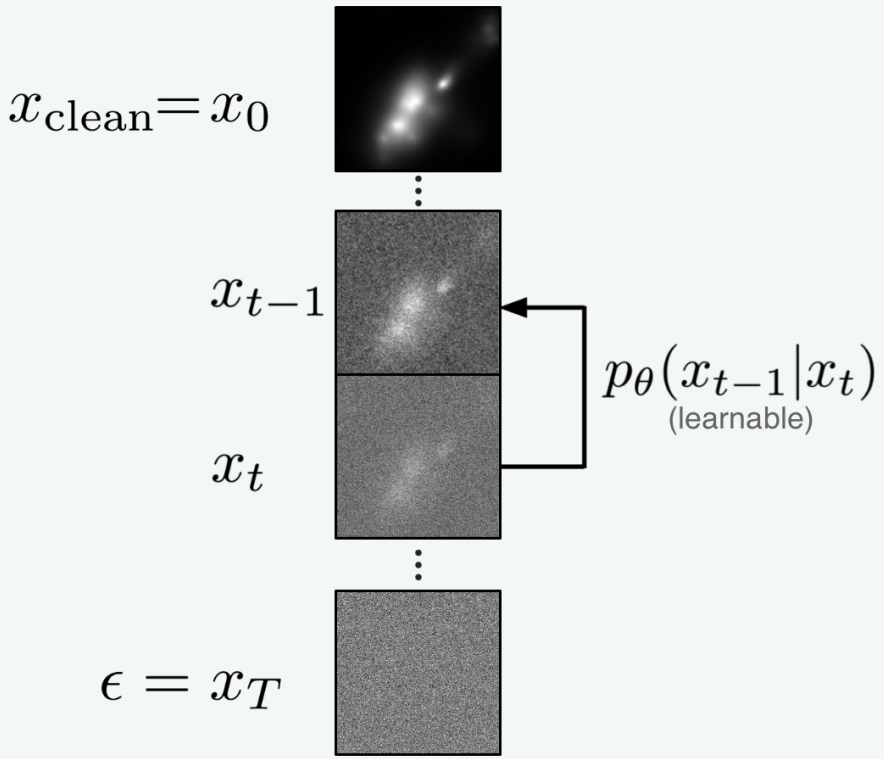
C. Bound is **tractable**: $q \sim \mathrm{N}(\mu(\epsilon), \cdot)$ and $p_\theta \sim \mathrm{N}(\mu(\epsilon_\theta), \cdot)$
   - Bayes' rule + Gaussian properties

D. Deduce loss function:
   - **Simple** regression-style noise prediction

Flavor 1:
Diffusion, DDPM

Generation Process
(must "reverse" the forward)

② Learn to **denoise** it



↬ STRATEGY ↫

A. Get (variational) **lower bound** on log-likelihood
   ↬ Jensen inequality ↭ E(vidence)L(ower)BO(und)

B. Expand terms of lower bound
   ↬ $\mathrm{KL}\left(q(x_{t-1}|x_t, x_0) \,\middle\|\, p_\theta(x_{t-1}|x_t)\right)$ pops up

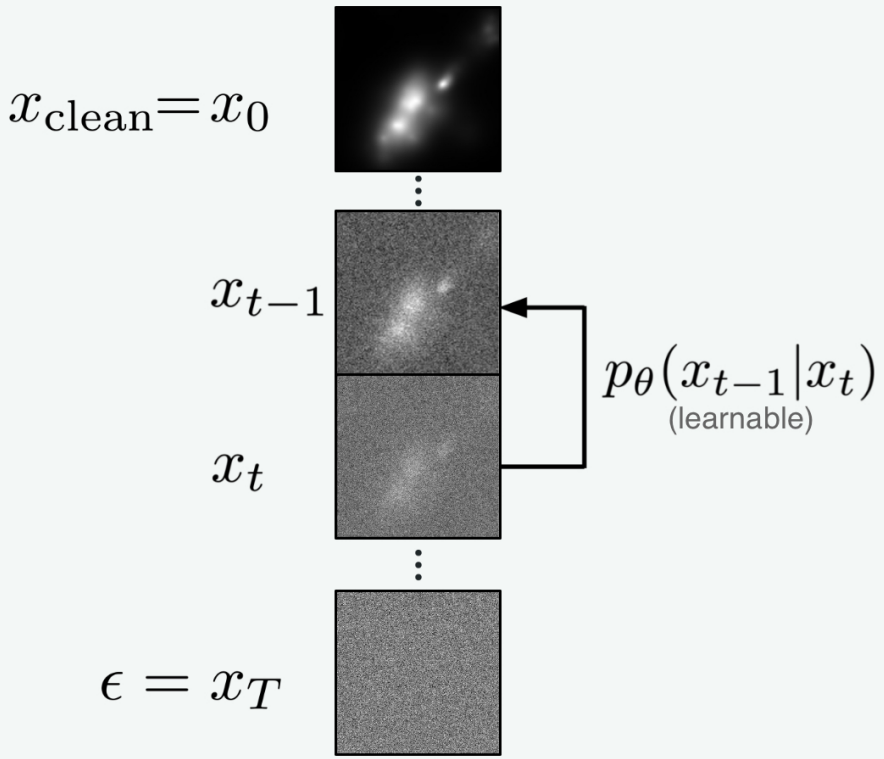
C. Bound is **tractable**: $q \sim \mathrm{N}(\mu(\epsilon), \cdot)$ and $p_\theta \sim \mathrm{N}(\mu(\epsilon_\theta), \cdot)$
   ↬ Bayes' rule + Gaussian properties

D. $$\mathcal{L}_{\text{DDPM}} = \mathbb{E}_{t,x_0,\epsilon}\left[\left\|\epsilon_\theta\left(\sqrt{\bar{\alpha}_t}\, x_0 + \sqrt{1-\bar{\alpha}_t}\,\epsilon,\, t\right) - \epsilon\right\|^2\right]$$

Flavor 1:
Diffusion, DDPM

Generation Process
(must "reverse" the forward)

② Learn to **denoise** it



↬ PROBLEM ↫

- As is, DDPM is too slow at inference time
  - ↬ Generation process is $O(T)$: ~minutes per sample

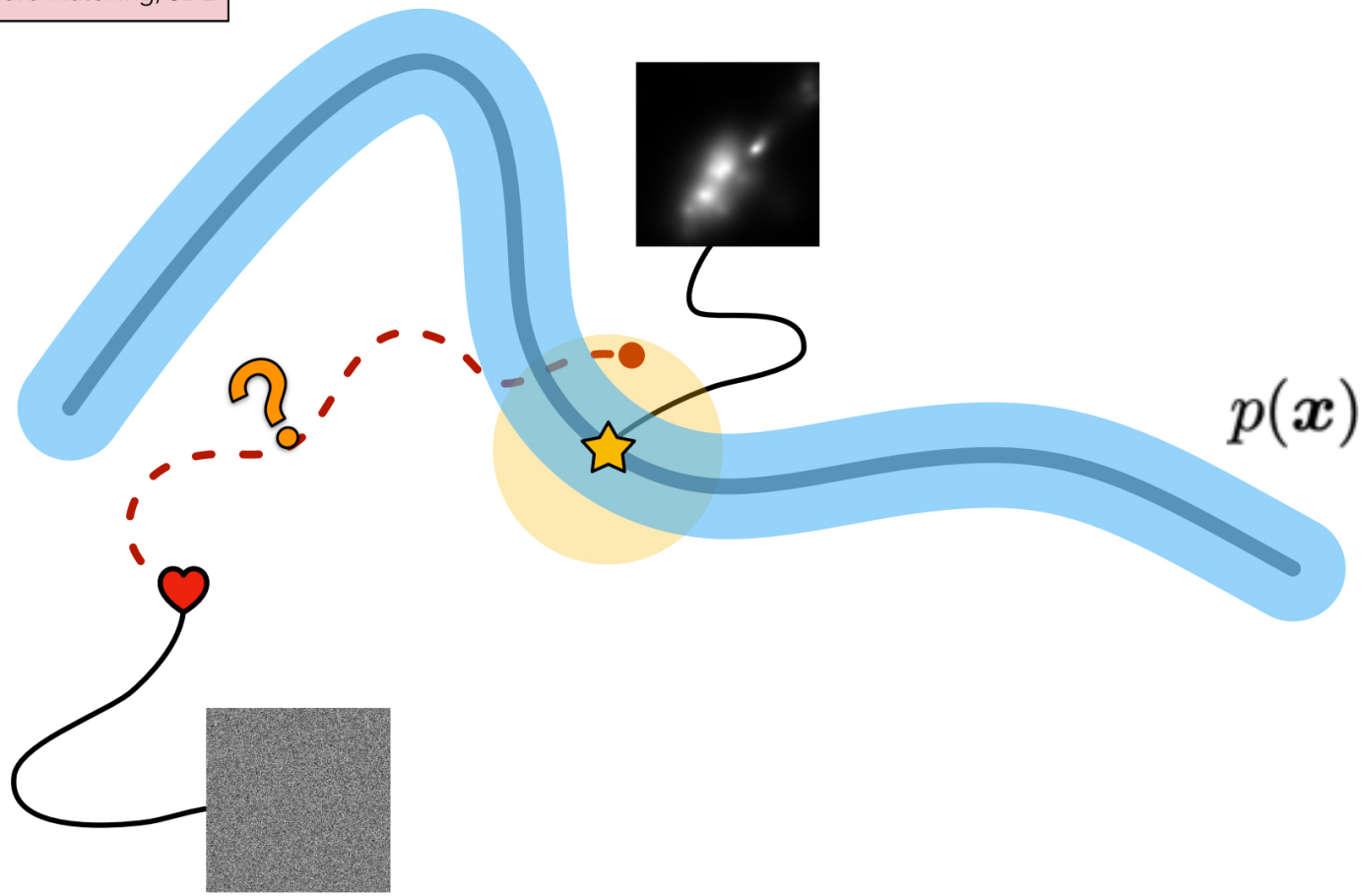
Published as a conference paper at ICLR 2021

DENOISING DIFFUSION IMPLICIT MODELS

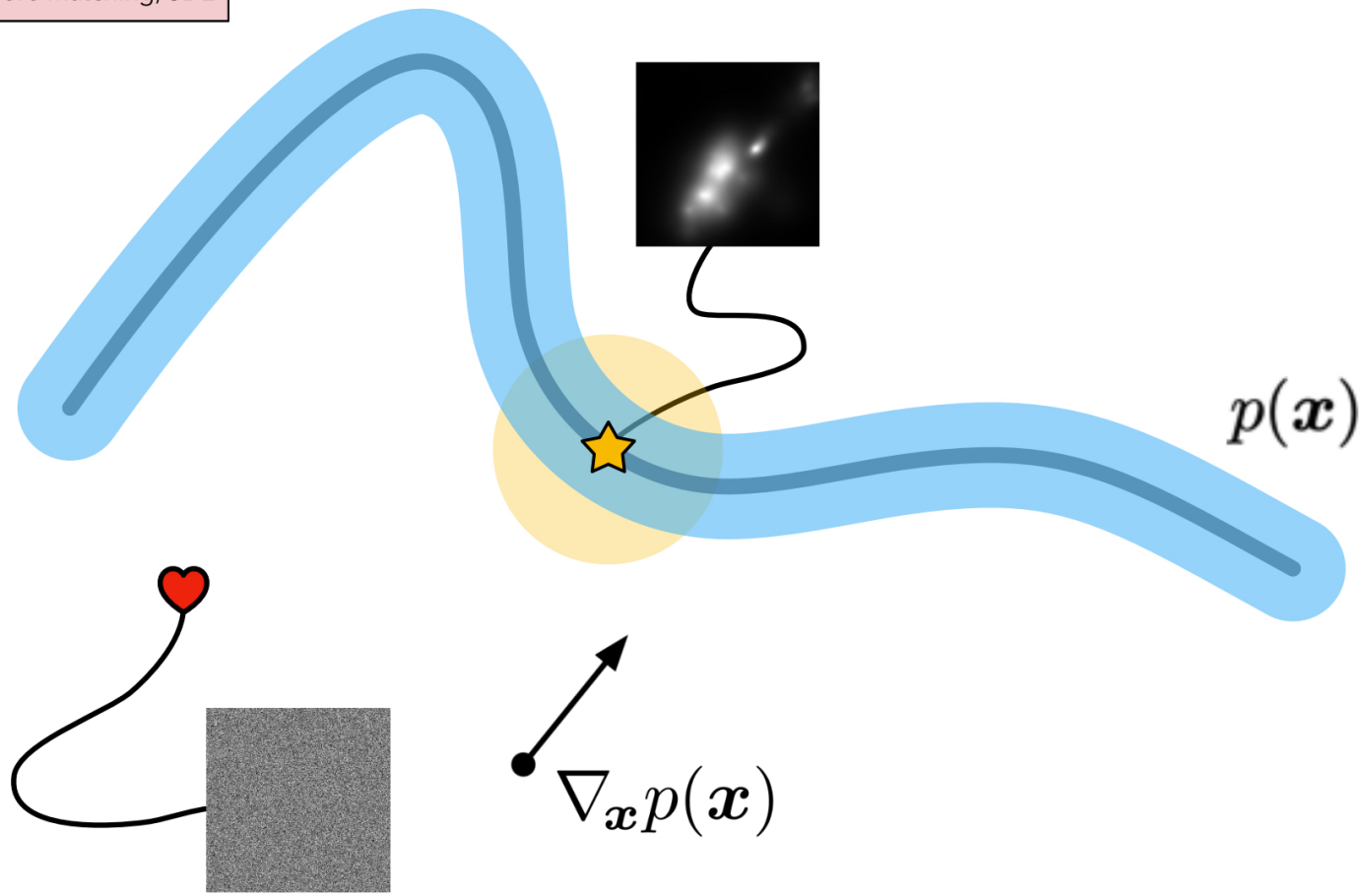
Jiaming Song, Chenlin Meng & Stefano Ermon
Stanford University
{tsong,chenlin,ermon}@cs.stanford.edu

A. Convenient assumptions compatible with DDPM optimization: same marginal but deterministic generation
B. No inter-step stochasticity
C. Skip steps

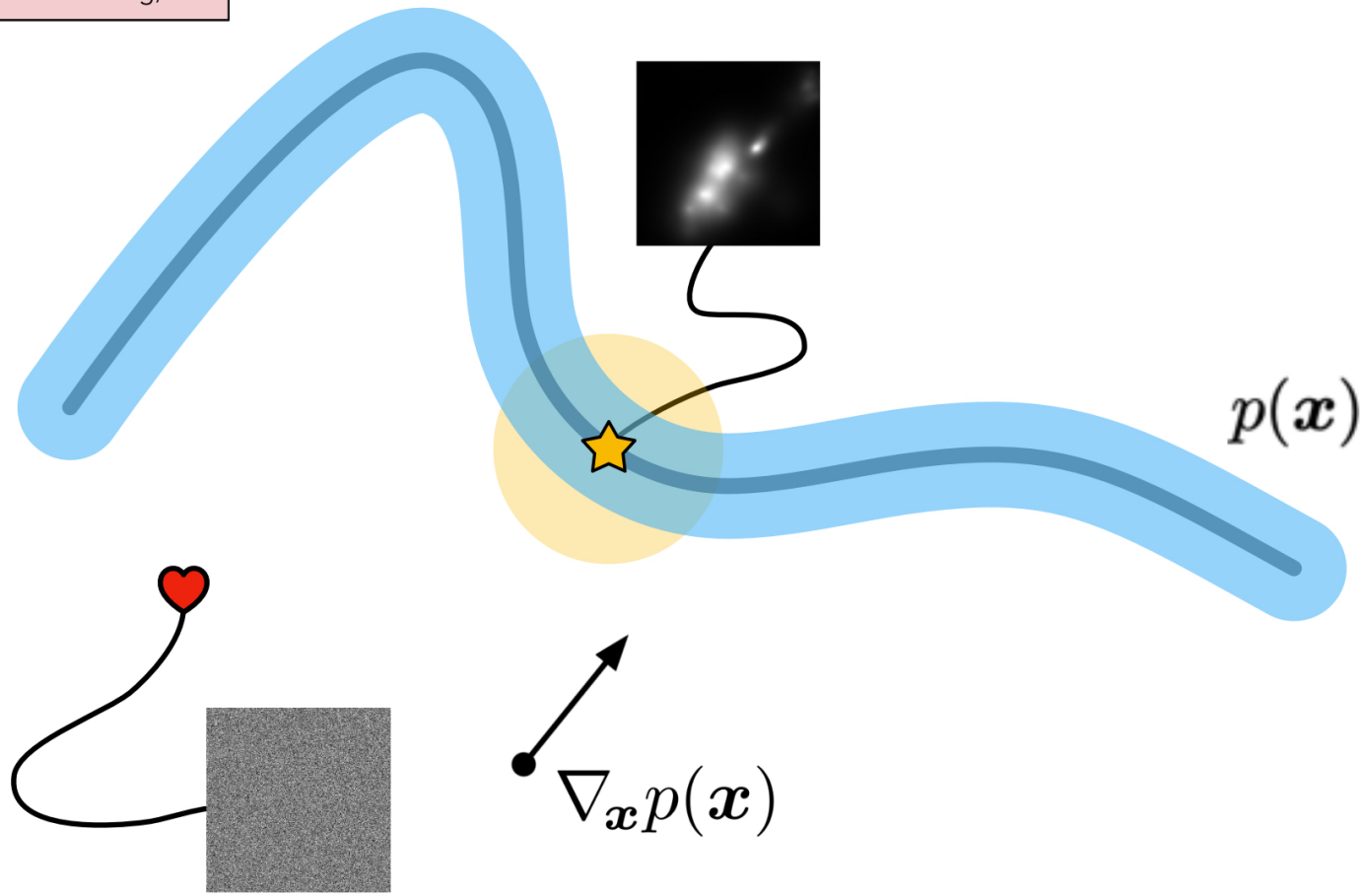
Flavor 2:
Score Matching, SDE



Flavor 2:
Score Matching, SDE



Flavor 2:
Score Matching, SDE



PROBLEMS

A. Typically the normalising constant of the data distribution is **intractable**

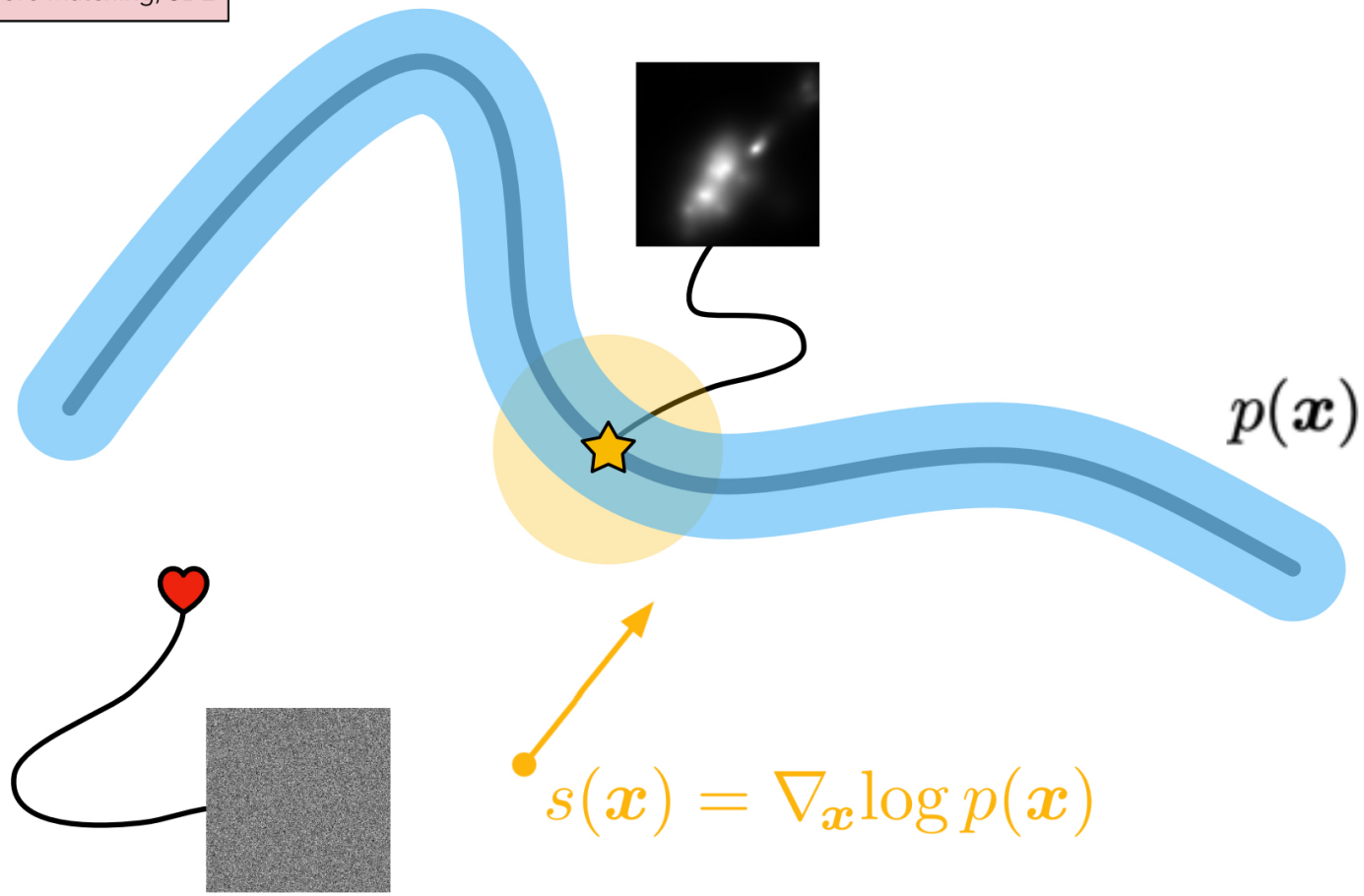
$$p(\boldsymbol{x}) = \frac{1}{Z} g(\boldsymbol{x}) \quad \text{with} \quad Z = \int g(\boldsymbol{x}) \mathrm{d}\boldsymbol{x}$$

B. Switching to gradient does **not** help!

$$\nabla_{\boldsymbol{x}} p(\boldsymbol{x}) = \nabla_{\boldsymbol{x}} \left(\frac{g(\boldsymbol{x})}{Z} \right)$$

C. Also, **numerically unstable** in low-density regions

Flavor 2:
Score Matching, SDE



SOLUTION

A. Go with the the **log-distribution**!

B. Its gradient is now **tractable**

$$\nabla_{\boldsymbol{x}} \log p(\boldsymbol{x}) = \nabla_{\boldsymbol{x}} \left(\log \frac{g(\boldsymbol{x})}{Z} \right) = \nabla_{\boldsymbol{x}} \log g(\boldsymbol{x})$$

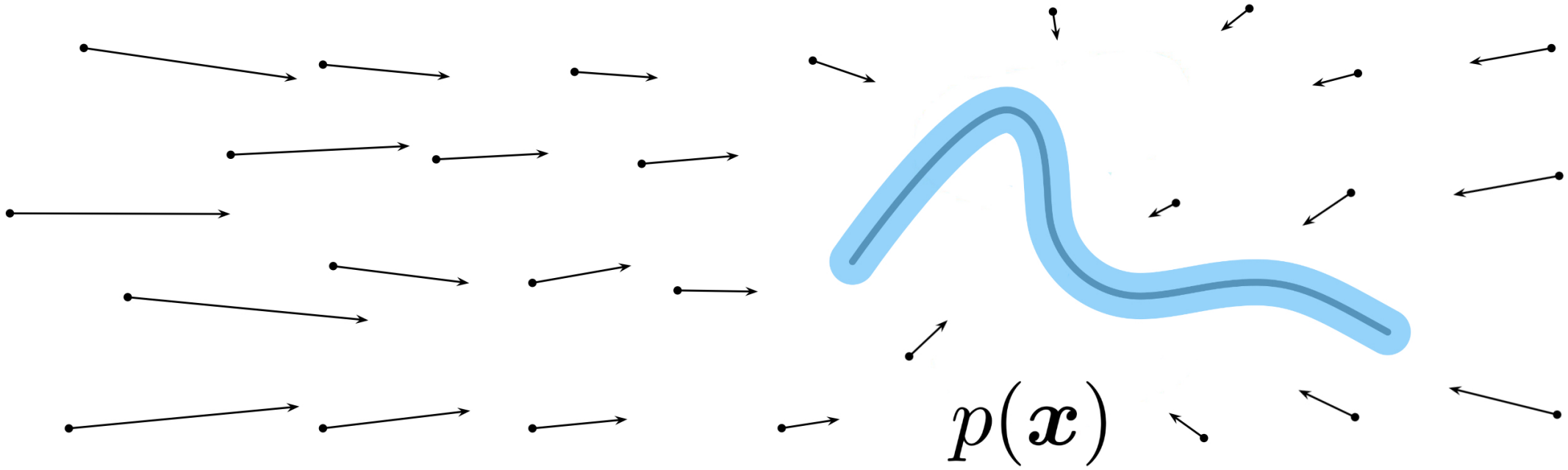
C. Same direction as the plain gradient

$$\nabla_{\boldsymbol{x}} \log p(\boldsymbol{x}) = \frac{\nabla_{\boldsymbol{x}} p(\boldsymbol{x})}{p(\boldsymbol{x})}$$

D. Also more **numerically stable**

Flavor 2:
Score Matching, SDE

$$s(\boldsymbol{x}) = \nabla_{\boldsymbol{x}} \log p(\boldsymbol{x})$$

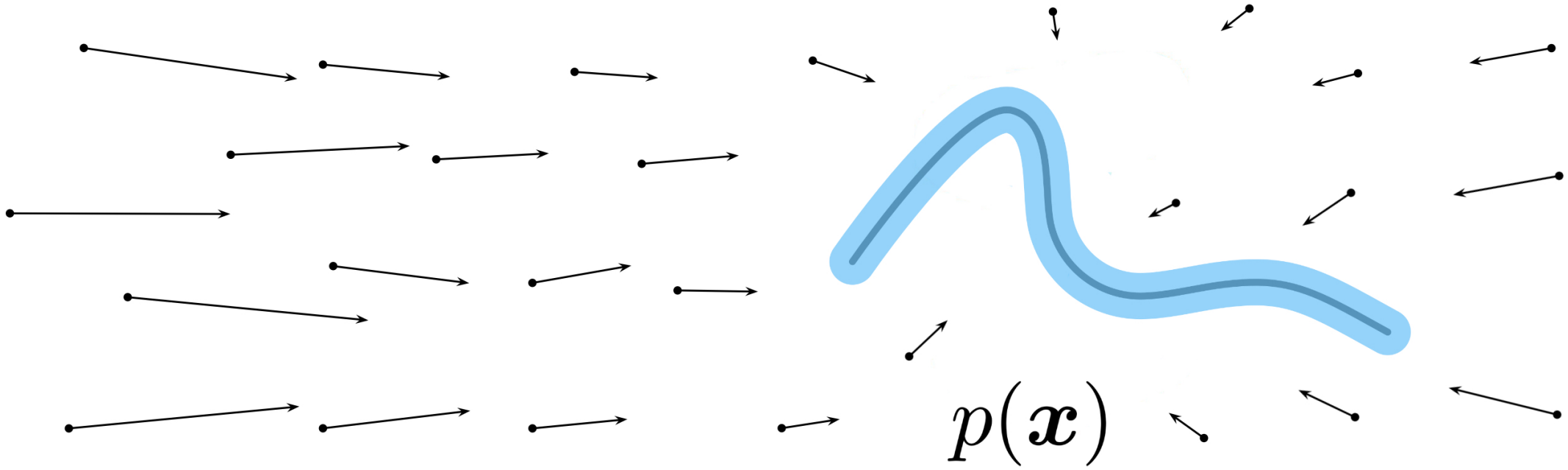


Flavor 2:
Score Matching, SDE

$$\mathcal{L}_{\mathrm{SM}} = \mathbb{E}_{\boldsymbol{x}}\left[\left\| s_\theta(\boldsymbol{x}) - \nabla_{\boldsymbol{x}} \log p(\boldsymbol{x}) \right\|^2\right]$$

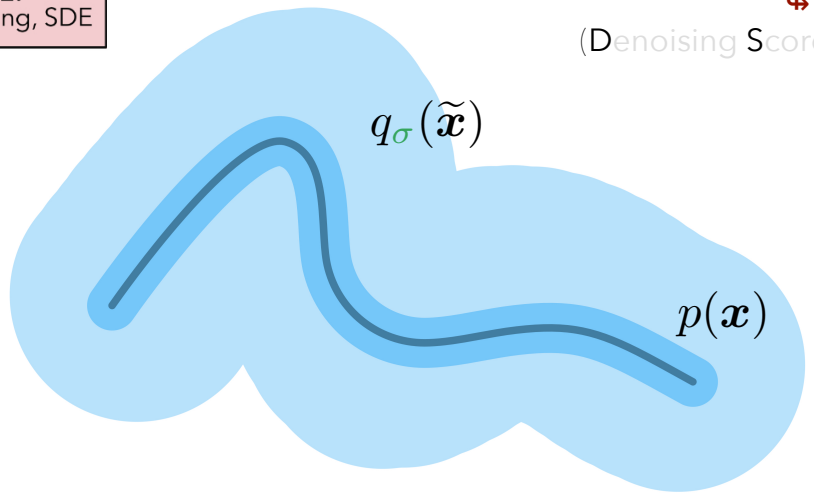
unknown!

↬ SCORE MATCHING ↫



Flavor 2:
Score Matching, SDE

IDEA 1/2
(Denoising Score Matching)



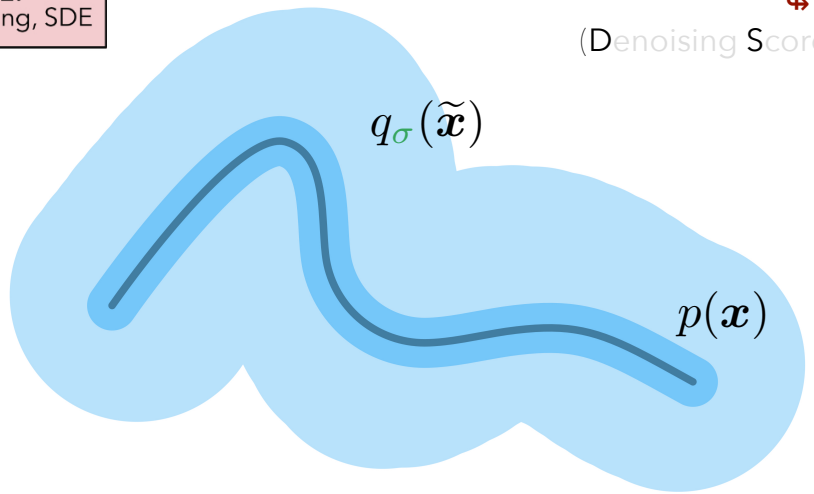
Add noise to data + leverage score closed form:

$$\widetilde{\boldsymbol{x}} = \boldsymbol{x} + \sigma\epsilon, \quad \sigma \gg 1, \quad \epsilon \sim \mathrm{N}(0, \mathbb{I})$$

$$q_\sigma(\widetilde{\boldsymbol{x}}|\boldsymbol{x}) = \mathrm{N}(\boldsymbol{x}, \sigma^2\mathbb{I}) \rightsquigarrow \nabla_{\widetilde{\boldsymbol{x}}} \log q_\sigma(\widetilde{\boldsymbol{x}}|\boldsymbol{x}) = -\frac{\widetilde{\boldsymbol{x}} - \boldsymbol{x}}{\sigma^2}$$

$$q_\sigma(\widetilde{\boldsymbol{x}}) = \int p(\boldsymbol{x}) q_\sigma(\widetilde{\boldsymbol{x}}|\boldsymbol{x}) \mathrm{d}\boldsymbol{x} = (p \star q_\sigma)(\widetilde{\boldsymbol{x}})$$

Flavor 2:
Score Matching, SDE



IDEA 1/2
(Denoising Score Matching)

Add noise to data + leverage score closed form:

$$\widetilde{\boldsymbol{x}} = \boldsymbol{x} + \sigma\epsilon, \quad \sigma \gg 1, \quad \epsilon \sim \mathrm{N}(0, \mathbb{I})$$

$$q_\sigma(\widetilde{\boldsymbol{x}}|\boldsymbol{x}) = \mathrm{N}(\boldsymbol{x}, \sigma^2\mathbb{I}) \;\rightsquigarrow\; \nabla_{\widetilde{\boldsymbol{x}}} \log q_\sigma(\widetilde{\boldsymbol{x}}|\boldsymbol{x}) = -\frac{\widetilde{\boldsymbol{x}} - \boldsymbol{x}}{\sigma^2}$$

$$q_\sigma(\widetilde{\boldsymbol{x}}) = \int p(\boldsymbol{x}) q_\sigma(\widetilde{\boldsymbol{x}}|\boldsymbol{x}) \mathrm{d}\boldsymbol{x} = (p \star q_\sigma)(\widetilde{\boldsymbol{x}})$$

CONSEQUENCE

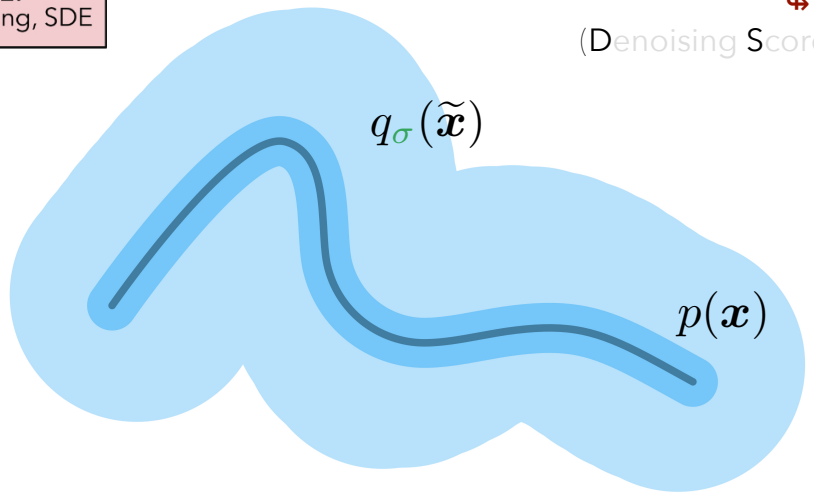
The new "smoothed" but complex loss

$$\mathcal{L}_{\mathrm{SM}}(q_\sigma) = \mathbb{E}_{\widetilde{\boldsymbol{x}}}\left[\left\| s_\theta(\widetilde{\boldsymbol{x}}) - \nabla_{\widetilde{\boldsymbol{x}}} \log q_\sigma(\widetilde{\boldsymbol{x}}) \right\|^2\right]$$

equals the tractable Denoising Score Matching loss
(open the square + def. of marginal)

$$\mathcal{L}_{\mathrm{DSM}} = \mathbb{E}_{\boldsymbol{x},\widetilde{\boldsymbol{x}}}\left[\left\| s_\theta(\widetilde{\boldsymbol{x}}) - \nabla_{\widetilde{\boldsymbol{x}}} \log q_\sigma(\widetilde{\boldsymbol{x}}|\boldsymbol{x}) \right\|^2\right]$$

Flavor 2:
Score Matching, SDE



IDEA 1/2
(Denoising Score Matching)

Add noise to data + leverage score closed form:

$$\tilde{\boldsymbol{x}} = \boldsymbol{x} + \sigma\epsilon, \quad \sigma \gg 1, \quad \epsilon \sim \mathrm{N}(0, \mathbb{I})$$

$$q_\sigma(\tilde{\boldsymbol{x}}|\boldsymbol{x}) = \mathrm{N}(\boldsymbol{x}, \sigma^2\mathbb{I}) \;\rightsquigarrow\; \nabla_{\tilde{\boldsymbol{x}}} \log q_\sigma(\tilde{\boldsymbol{x}}|\boldsymbol{x}) = -\frac{\tilde{\boldsymbol{x}} - \boldsymbol{x}}{\sigma^2}$$

$$q_\sigma(\tilde{\boldsymbol{x}}) = \int p(\boldsymbol{x}) q_\sigma(\tilde{\boldsymbol{x}}|\boldsymbol{x}) \mathrm{d}\boldsymbol{x} = (p \star q_\sigma)(\tilde{\boldsymbol{x}})$$

CONSEQUENCE

The new "smoothed" but complex loss

$$\mathcal{L}_{\mathrm{SM}}(q_\sigma) = \mathbb{E}_{\tilde{\boldsymbol{x}}}\left[\left\| s_\theta(\tilde{\boldsymbol{x}}) - \nabla_{\tilde{\boldsymbol{x}}} \log q_\sigma(\tilde{\boldsymbol{x}}) \right\|^2\right]$$

equals the tractable Denoising Score Matching loss
(open the square + def. of marginal)

$$\mathcal{L}_{\mathrm{DSM}} = \mathbb{E}_{\boldsymbol{x},\tilde{\boldsymbol{x}}}\left[\left\| s_\theta(\tilde{\boldsymbol{x}}) - \nabla_{\tilde{\boldsymbol{x}}} \log q_\sigma(\tilde{\boldsymbol{x}}|\boldsymbol{x}) \right\|^2\right]$$

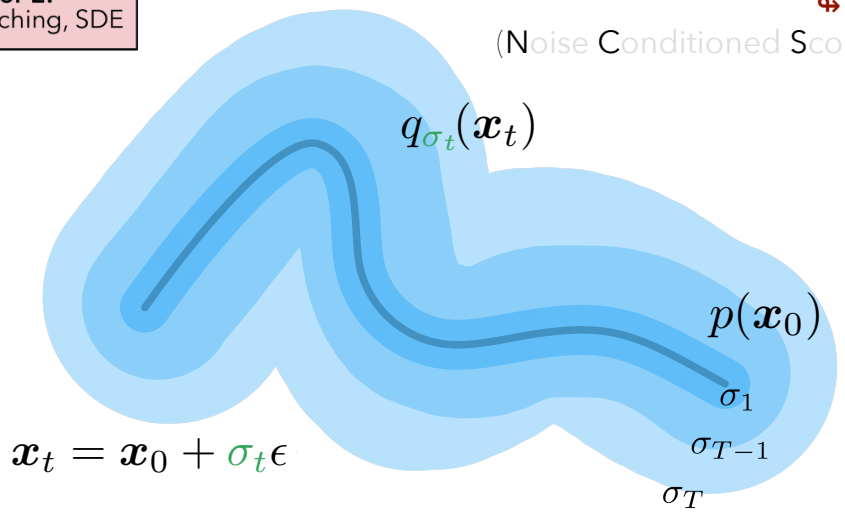
PROBLEM

Good in low-density regions but **wrong target!**

$$\sigma \gg 1 \quad \rightsquigarrow \quad p(\cdot) \neq q_\sigma(\cdot)$$

Flavor 2:
Score Matching, SDE

↬ IDEA 2/2 ↫
(Noise Conditioned Score Network)



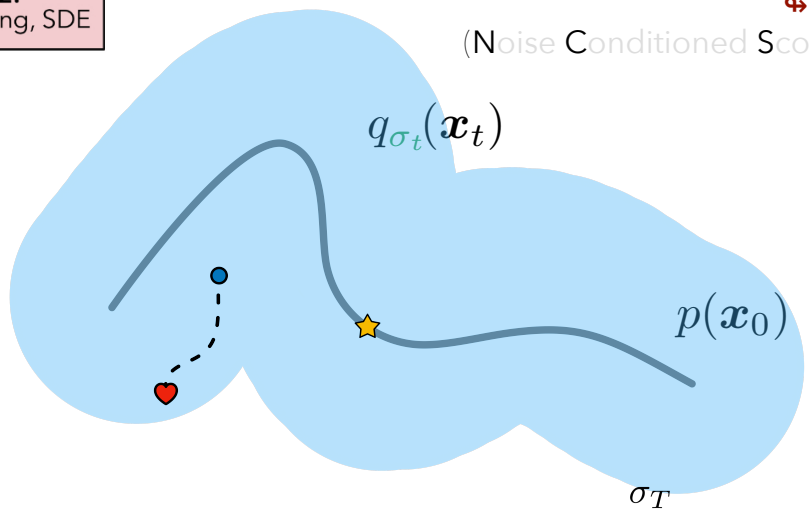
- Estimate score at different levels $\sigma_T > \cdots > \sigma_1$

$$\mathcal{L}_{\text{NCSN}} = \frac{1}{T} \sum_{t=1}^{T} \lambda_t \, \mathbb{E}_{\boldsymbol{x}_t} \left[\left\| s_\theta(\boldsymbol{x}_t, \sigma_t) - \nabla_{\boldsymbol{x}_t} \log q_{\sigma_t}(\boldsymbol{x}_t) \right\|^2 \right]$$

- Essentially an average over "time" of $\mathcal{L}_{\text{DSM}}$ losses
- The weights λ_t compensate for the fact that scores explode as σ_t decreases: $-\nabla_{\boldsymbol{x}_t} \log q_{\sigma_t}(\boldsymbol{x}_t | \boldsymbol{x}_0) \propto 1/\sigma_t^2$

Flavor 2:
Score Matching, SDE

IDEA 2/2
(Noise Conditioned Score Network)

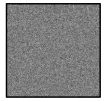


- Estimate score at different levels $\sigma_T > \cdots > \sigma_1$

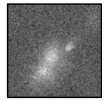
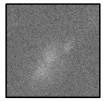
$$\mathcal{L}_{\text{NCSN}} = \frac{1}{T}\sum_{t=1}^{T} \lambda_t \, \mathbb{E}_{\boldsymbol{x}_t}\Big[\big\| s_\theta(\boldsymbol{x}_t, \sigma_t) - \nabla_{\boldsymbol{x}_t} \log q_{\sigma_t}(\boldsymbol{x}_t)\big\|^2\Big]$$

- Essentially an average over "time" of $\mathcal{L}_{\text{DSM}}$ losses
- The weights λ_t compensate for the fact that scores explode as σ_t decreases: $-\nabla_{\boldsymbol{x}_t} \log q_{\sigma_t}(\boldsymbol{x}_t|\boldsymbol{x}_0) \propto 1/\sigma_t^2$

GENERATION | "Backward"
(Annealed Langevin Dynamics)

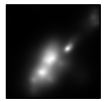
① Sample: $\boldsymbol{x}_T =$  $\sim \mathrm{N}(0, \sigma_T \mathbb{I})$

② For each $\sigma_T > \cdots > \sigma_1$ perform K Langevin updates:

$$\text{} \leftarrow \text{} + \frac{\alpha_t}{2} s_\theta(\text{}, \sigma_t) + \sqrt{\alpha_t}\,\boldsymbol{\eta}$$

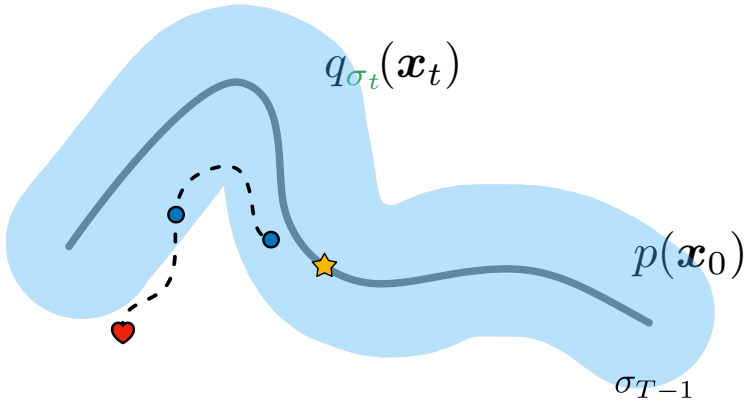
Predicted Score (direction) — **Sampled** from a standard Normal

③ Get the final image:

$\boldsymbol{x}_0 =$ 

Flavor 2:
Score Matching, SDE

IDEA 2/2
(Noise Conditioned Score Network)

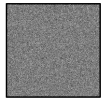


- Estimate score at different levels $\sigma_T > \cdots > \sigma_1$

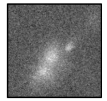
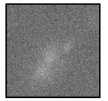
$$\mathcal{L}_{\text{NCSN}} = \frac{1}{T}\sum_{t=1}^{T} \lambda_t \, \mathbb{E}_{\boldsymbol{x}_t}\Big[\big\| s_\theta(\boldsymbol{x}_t, \sigma_t) - \nabla_{\boldsymbol{x}_t} \log q_{\sigma_t}(\boldsymbol{x}_t)\big\|^2\Big]$$

- Essentially an average over "time" of $\mathcal{L}_{\text{DSM}}$ losses
- The weights λ_t compensate for the fact that scores explode as σ_t decreases: $-\nabla_{\boldsymbol{x}_t} \log q_{\sigma_t}(\boldsymbol{x}_t|\boldsymbol{x}_0) \propto 1/\sigma_t^2$

GENERATION | "Backward"
(Annealed Langevin Dynamics)

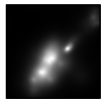
① Sample: $\boldsymbol{x}_T =$  $\sim \mathrm{N}(0, \sigma_T \mathbb{I})$

② For each $\sigma_T > \cdots > \sigma_1$ perform K Langevin updates:

$$\text{} \leftarrow \text{} + \frac{\alpha_t}{2} s_\theta(\text{}, \sigma_t) + \sqrt{\alpha_t}\,\boldsymbol{\eta}$$

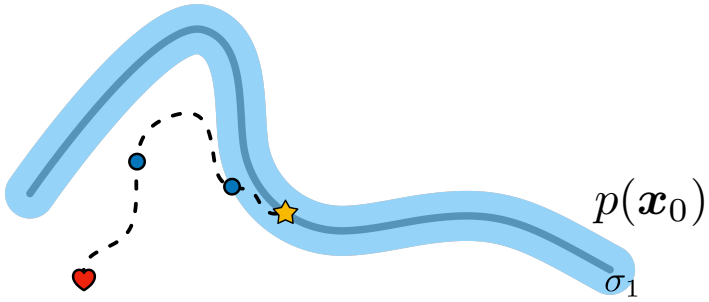
Predicted Score (direction) — **Sampled** from a standard Normal

③ Get the final image:

$\boldsymbol{x}_0 =$ 

Flavor 2:
Score Matching, SDE

IDEA 2/2
(Noise Conditioned Score Network)

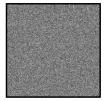


- Estimate score at different levels $\sigma_T > \cdots > \sigma_1$

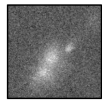
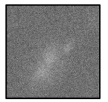
$$\mathcal{L}_{\text{NCSN}} = \frac{1}{T}\sum_{t=1}^{T} \lambda_t \, \mathbb{E}_{\boldsymbol{x}_t}\Big[\big\| s_\theta(\boldsymbol{x}_t, \sigma_t) - \nabla_{\boldsymbol{x}_t} \log q_{\sigma_t}(\boldsymbol{x}_t)\big\|^2\Big]$$

- Essentially an average over "time" of $\mathcal{L}_{\text{DSM}}$ losses
- The weights λ_t compensate for the fact that scores explode as σ_t decreases: $-\nabla_{\boldsymbol{x}_t} \log q_{\sigma_t}(\boldsymbol{x}_t|\boldsymbol{x}_0) \propto 1/\sigma_t^2$

GENERATION | "Backward"
(Annealed Langevin Dynamics)

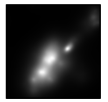
① Sample: $\boldsymbol{x}_T =$  $\sim \mathrm{N}(0, \sigma_T \mathbb{I})$

② For each $\sigma_T > \cdots > \sigma_1$ perform K Langevin updates:

 $\leftarrow$  $+ \frac{\alpha_t}{2} s_\theta($ $, \sigma_t) + \sqrt{\alpha_t}\,\boldsymbol{\eta}$

Predicted Score (direction) — **Sampled** from a standard Normal

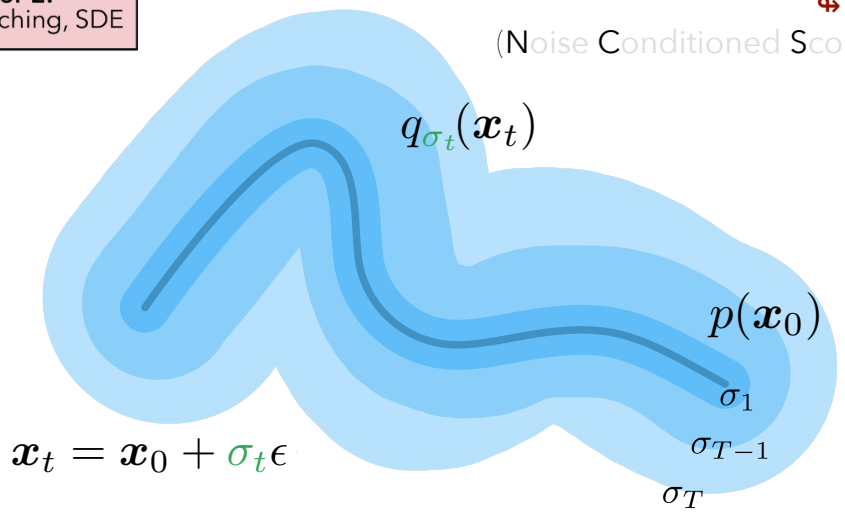
③ Get the final image:

$\boldsymbol{x}_0 =$ 

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
INAF ISTITUTO NAZIONALE DI ASTROFISICA

Flavor 2:
Score Matching, SDE

IDEA 2/2
(Noise Conditioned Score Network)

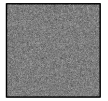


- Estimate score at different levels $\sigma_T > \cdots > \sigma_1$

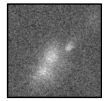
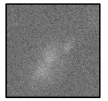
$$\mathcal{L}_{\text{NCSN}} = \frac{1}{T}\sum_{t=1}^{T} \lambda_t \, \mathbb{E}_{\boldsymbol{x}_t}\Big[\big\| s_\theta(\boldsymbol{x}_t, \sigma_t) - \nabla_{\boldsymbol{x}_t} \log q_{\sigma_t}(\boldsymbol{x}_t)\big\|^2\Big]$$

- Essentially an average over "time" of $\mathcal{L}_{\text{DSM}}$ losses
- The weights λ_t compensate for the fact that scores explode as σ_t decreases: $-\nabla_{\boldsymbol{x}_t} \log q_{\sigma_t}(\boldsymbol{x}_t|\boldsymbol{x}_0) \propto 1/\sigma_t^2$

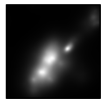
GENERATION | "Backward"
(Annealed Langevin Dynamics)

① Sample: $\boldsymbol{x}_T =$  $\sim \mathrm{N}(0, \sigma_T \mathbb{I})$

② For each $\sigma_T > \cdots > \sigma_1$ perform K Langevin updates:

$$\text{} \leftarrow \text{} + \frac{\alpha_t}{2} s_\theta(\text{}, \sigma_t) + \sqrt{\alpha_t}\,\boldsymbol{\eta}$$

Predicted Score (direction) — **Sampled** from a standard Normal

③ Get the final image: $\boldsymbol{x}_0 =$ 

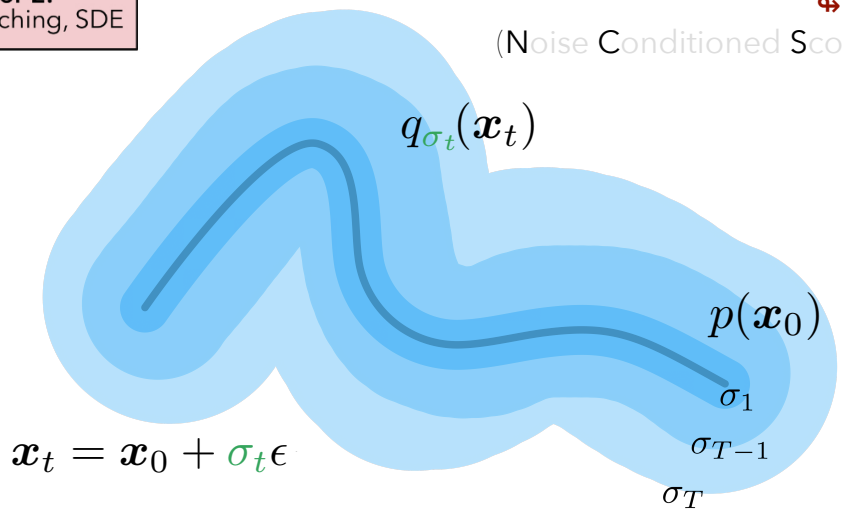
NOTICE | "Forward"

DDPM | Variance Preserving		NCSN | Variance Exploding
$\boldsymbol{x}_t = \sqrt{\bar{\alpha}_t} \cdot \boldsymbol{x}_0 + \sqrt{1-\bar{\alpha}_t} \cdot \epsilon$		$\boldsymbol{x}_t = 1 \cdot \boldsymbol{x}_0 + \sigma_t \cdot \epsilon$
$\nabla_{\boldsymbol{x}_t} \log q(\boldsymbol{x}_t\|\boldsymbol{x}_0) = -\epsilon/\sqrt{1-\bar{\alpha}_t}$	vs	$\nabla_{\boldsymbol{x}_t} \log q(\boldsymbol{x}_t\|\boldsymbol{x}_0) = -\epsilon/\sigma_t$
signal-Var ↓ noise-Var ↑ tot-Var ≈ c		signal-Var ≈ c noise-Var ↑ tot-Var ↑

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
INAF ISTITUTO NAZIONALE DI ASTROFISICA

Flavor 2:
Score Matching, SDE

IDEA 2/2
(Noise Conditioned Score Network)

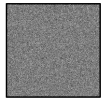


- Estimate score at different levels $\sigma_T > \cdots > \sigma_1$

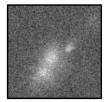
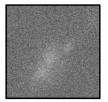
$$\mathcal{L}_{\text{NCSN}} = \frac{1}{T}\sum_{t=1}^{T} \lambda_t \, \mathbb{E}_{\boldsymbol{x}_t}\left[\left\| s_\theta(\boldsymbol{x}_t, \sigma_t) - \nabla_{\boldsymbol{x}_t} \log q_{\sigma_t}(\boldsymbol{x}_t)\right\|^2\right]$$

- Essentially an average over "time" of $\mathcal{L}_{\text{DSM}}$ losses
- The weights λ_t compensate for the fact that scores explode as σ_t decreases: $-\nabla_{\boldsymbol{x}_t} \log q_{\sigma_t}(\boldsymbol{x}_t|\boldsymbol{x}_0) \propto 1/\sigma_t^2$

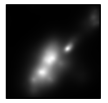
GENERATION | "Backward"
(Annealed Langevin Dynamics)

① Sample: $\boldsymbol{x}_T =$  $\sim \mathrm{N}(0, \sigma_T \mathbb{I})$

② For each $\sigma_T > \cdots > \sigma_1$ perform K Langevin updates:

 $\leftarrow$  $+ \frac{\alpha_t}{2} s_\theta($ $, \sigma_t) + \sqrt{\alpha_t}\,\boldsymbol{\eta}$

Predicted Score (direction) — Sampled from a standard Normal

③ Get the final image: $\boldsymbol{x}_0 =$ 

NOTICE | Hence…

$$\mathcal{L}_{\text{DDPM}} \equiv \mathcal{L}_{\text{NCSN}}$$

PNRR Missione 4 • Componente 2 Investimento 3.1
STILES – IR0000034 CUP C33C22000640006

. . . uhm, what if we consider **continuous** evolutions?



. . . uhm, what if we consider **continuous** evolutions?



…well, we “just” need a continuous counterpart to add increments of Gaussian noise! 🤓

Flavor 2:
Score Matching, SDE

Forward Process

DDPM

Discrete

$$x_t = \sqrt{1-\beta_t}\, x_{t-1} + \sqrt{\beta_t}\,\epsilon$$

$\Delta t \downarrow 0$

Continuous

$$dx = -\frac{1}{2}\beta(t)\, x\, dt + \sqrt{\beta(t)}\, dW$$

NCSN

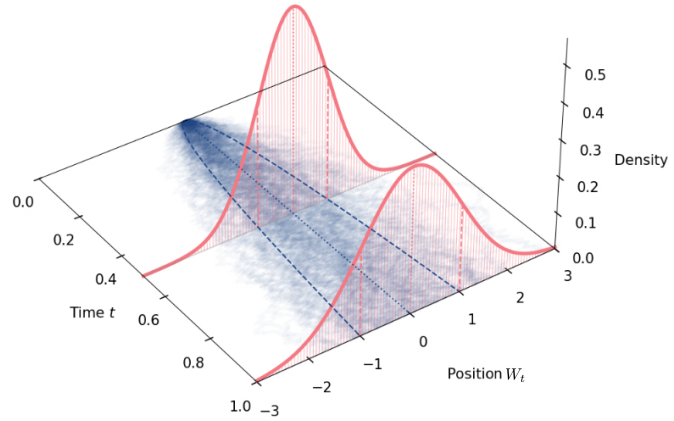
Discrete

$$x_t = x_{t-1} + \sqrt{\sigma_t^2 - \sigma_{t-1}^2}\epsilon$$

$\Delta t \downarrow 0$

Continuous

$$dx = \sqrt{\frac{d[\sigma^2(t)]}{dt}}\, dW$$



Wiener Process or Brownian Motion

- $W_0 = 0$
- $W_t - W_s \sim \mathrm{N}(0, (t-s)\mathbb{I})$
- Independent Increments

Flavor 2:
Score Matching, SDE

Forward Process

DDPM

Discrete

$$x_t = \sqrt{1-\beta_t}\, x_{t-1} + \sqrt{\beta_t}\,\epsilon$$

$\Delta t \downarrow 0$

Continuous

$$dx = -\frac{1}{2}\beta(t)\, x\, dt + \sqrt{\beta(t)}\, dW$$

NCSN

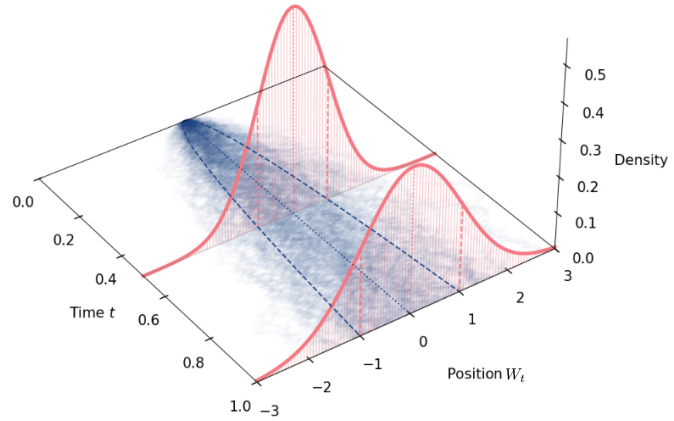
Discrete

$$x_t = x_{t-1} + \sqrt{\sigma_t^2 - \sigma_{t-1}^2}\epsilon$$

$\Delta t \downarrow 0$

Continuous

$$dx = \sqrt{\frac{d[\sigma^2(t)]}{dt}}\, dW$$



Wiener Process or Brownian Motion

- $W_0 = 0$
- $W_t - W_s \sim \mathrm{N}(0, (t-s)\mathbb{I})$
- Independent Increments

$$dx = \underbrace{f(x,t)dt}_{\texttt{drift}} + \underbrace{g(t)dW}_{\texttt{diffusion}}$$

Stochastic Differential Equation

Flavor 2:
Score Matching, SDE

Backward Process

MAIN RESULT

$$dx = \underbrace{f(x,t)dt}_{\texttt{drift}} + \underbrace{g(t)dW}_{\texttt{diffusion}}$$



The Annals of Probability
1986, Vol. 14, No. 4, 1188–1205

TIME REVERSAL OF DIFFUSIONS[1]

BY U. G. HAUSSMANN AND E. PARDOUX

University of British Columbia and Université de Provence

It is shown that if a diffusion process, $\{X_t: 0 \le t \le 1\}$, on R^d satisfies

$$dX_t = b(t, X_t)\,dt + \sigma(t, X_t)\,dw_t$$

then the reversed process, $\{\overline{X}_t: 0 \le t \le 1\}$ where $\overline{X}_t = X_{1-t}$, is again a diffusion with drift $\bar{b}$ and diffusion coefficient $\bar{\sigma}$, provided some mild conditions on b, σ, and p_0, the density of the law of X_0, hold. Moreover $\bar{b}$ and $\bar{\sigma}$ are identified.

$$\frac{\partial q_t(x)}{\partial t} = -\nabla_x\big(f(x,t)q_t(x)\big) + \frac{1}{2}g(t)^2\Delta_x q_t(x)$$
$$\texttt{Fokker}-\texttt{Planck Equation}$$

$$dx = \underbrace{\Big[f(x,t) - g(t)^2\nabla_x \log q_t(x)\Big]}_{\texttt{correction to drift}}\,dt + g(t)d\widetilde{W}$$

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
INAF ISTITUTO NAZIONALE DI ASTROFISICA

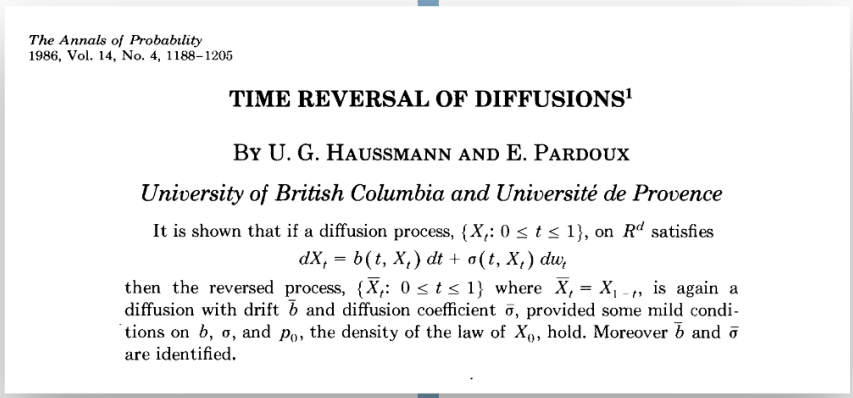
Flavor 2:
Score Matching, SDE

Backward Process

MAIN RESULT

$$dx = \underbrace{f(x,t)dt}_{\texttt{drift}} + \underbrace{g(t)dW}_{\texttt{diffusion}}$$

(deterministic) (stochastic)



The Annals of Probability
1986, Vol. 14, No. 4, 1188–1205

TIME REVERSAL OF DIFFUSIONS[1]

BY U. G. HAUSSMANN AND E. PARDOUX

University of British Columbia and Université de Provence

It is shown that if a diffusion process, $\{X_t\colon 0 \le t \le 1\}$, on R^d satisfies

$$dX_t = b(t, X_t)\,dt + \sigma(t, X_t)\,dw_t$$

then the reversed process, $\{\overline{X}_t\colon 0 \le t \le 1\}$ where $\overline{X}_t = X_{1-t}$, is again a diffusion with drift $\bar{b}$ and diffusion coefficient $\bar{\sigma}$, provided some mild conditions on b, σ, and p_0, the density of the law of X_0, hold. Moreover $\bar{b}$ and $\bar{\sigma}$ are identified.

Describe the evolution of the **marginal distribution** of the process
For the **Brownian motion** it is "just" the **heat equation**

$$\frac{\partial q_t(x)}{\partial t} = -\nabla_x\big(f(x,t)q_t(x)\big) + \frac{1}{2}g(t)^2\Delta_x q_t(x)$$

Fokker–Planck Equation

For **DDPM** and **NCSN**, we can solve it **analytically**
to get the convolution-type distribution we already saw

$$dx = \Big[f(x,t) - \underbrace{g(t)^2\nabla_x \log q_t(x)}_{\texttt{correction to drift}}\Big]\,dt + g(t)d\widetilde{W}$$

(..to be run backward in time)

The score!

PNRR Missione 4 • Componente 2 Investimento 3.1
STILES - IR0000034 CUP C33C22000640006

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
INAF ISTITUTO NAZIONALE DI ASTROFISICA

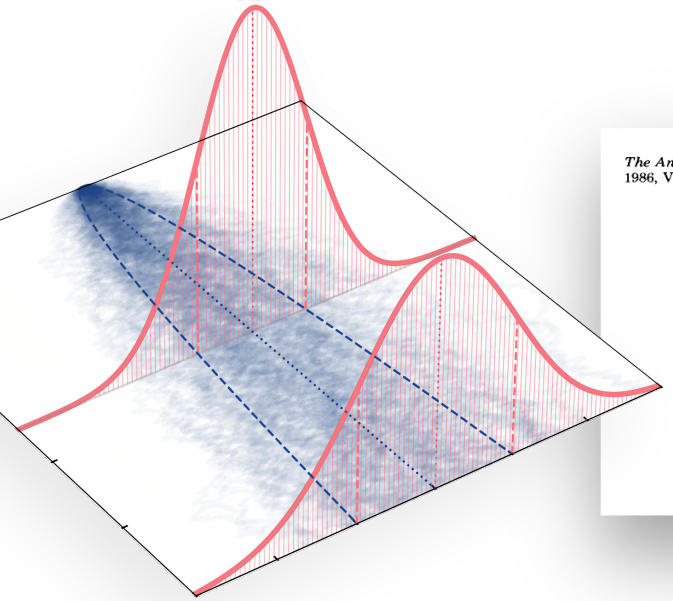
Flavor 2:
Score Matching, SDE

Backward Process

MAIN RESULT

$$dx = \underbrace{f(x,t)dt}_{\texttt{drift}} + \underbrace{g(t)dW}_{\texttt{diffusion}}$$

(deterministic) (stochastic)



The Annals of Probability
1986, Vol. 14, No. 4, 1188–1205

TIME REVERSAL OF DIFFUSIONS[1]

BY U. G. HAUSSMANN AND E. PARDOUX

University of British Columbia and Université de Provence

It is shown that if a diffusion process, $\{X_t\colon 0 \le t \le 1\}$, on R^d satisfies

$$dX_t = b(t, X_t)\,dt + \sigma(t, X_t)\,dw_t$$

then the reversed process, $\{\overline{X}_t\colon 0 \le t \le 1\}$ where $\overline{X}_t = X_{1-t}$, is again a diffusion with drift $\bar{b}$ and diffusion coefficient $\bar{\sigma}$, provided some mild conditions on b, σ, and p_0, the density of the law of X_0, hold. Moreover $\bar{b}$ and $\bar{\sigma}$ are identified.

Divergence measures if the transport produces a net outward/inward flow at a given location.

Laplacian move mass from high to low density regions

Change of Density = Drift-induced Transport + Diffusive Spreading

`Fokker—Planck Equation`

For **DDPM** and **NCSN**, we can solve it **analytically** to get the convolution-type distribution we already saw

$$dx = \Big[f(x,t) - \underbrace{g(t)^2 \nabla_x \log q_t(x)}_{\texttt{correction to drift}}\Big]\,dt + g(t)d\widetilde{W}$$

(..to be run backward in time)

The score!

PNRR Missione 4 • Componente 2 Investimento 3.1
STILES - IR0000034 CUP C33C22000640006

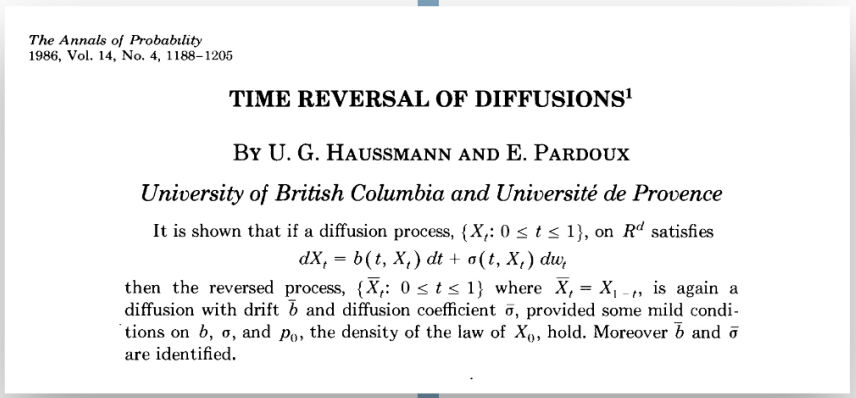
Flavor 2:
Score Matching, SDE

Backward Process

SOLVER

$$dx = f(x,t)dt + g(t)dW$$

drift diffusion



The Annals of Probability
1986, Vol. 14, No. 4, 1188–1205

TIME REVERSAL OF DIFFUSIONS[1]

BY U. G. HAUSSMANN AND E. PARDOUX

University of British Columbia and Université de Provence

It is shown that if a diffusion process, $\{X_t: 0 \le t \le 1\}$, on R^d satisfies

$$dX_t = b(t, X_t)\,dt + \sigma(t, X_t)\,dw_t$$

then the reversed process, $\{\overline{X}_t: 0 \le t \le 1\}$ where $\overline{X}_t = X_{1-t}$, is again a diffusion with drift $\bar{b}$ and diffusion coefficient $\bar{\sigma}$, provided some mild conditions on b, σ, and p_0, the density of the law of X_0, hold. Moreover $\bar{b}$ and $\bar{\sigma}$ are identified.

Describe the evolution of the **marginal distribution** of the process

$$\frac{\partial q_t(x)}{\partial t} = -\nabla_x\big(f(x,t)q_t(x)\big) + \frac{1}{2}g(t)^2\Delta_x q_t(x)$$

Fokker–Planck Equation

For **DDPM** and **NCSN**, we can solve it **analytically**
to get the convolution-type distribution we already saw

$$x_{t_{i-1}} = x_{t_i} + \left[f(x_{t_i}, t_i) - g(t_i)^2 s_\theta(x_{t_i}, t_i)\right]\Delta t + g(t_i)\sqrt{\Delta t}\,\xi_i$$

Euler–Maruyama Discretization

$\sim \mathrm{N}(0, \mathbb{I})$

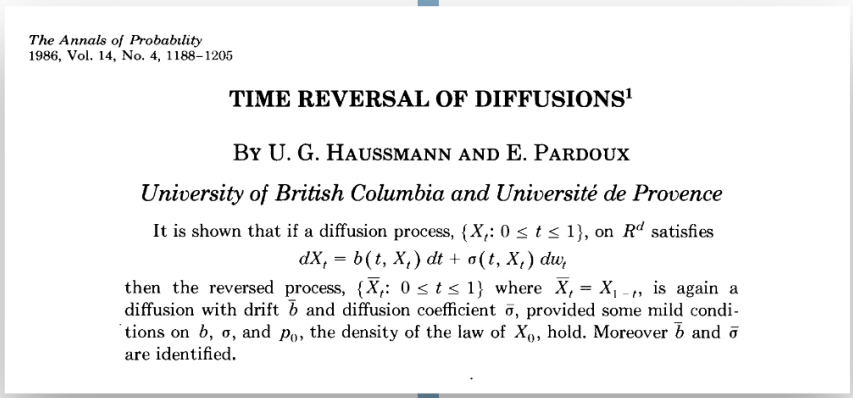
(...SDE solver for inference)

Flavor 2:
Score Matching, SDE

Backward Process

$$dx = \underbrace{f(x,t)dt}_{\texttt{drift}} + \underbrace{g(t)dW}_{\texttt{diffusion}}$$

PROBLEMS



The Annals of Probability
1986, Vol. 14, No. 4, 1188–1205

TIME REVERSAL OF DIFFUSIONS[1]

BY U. G. HAUSSMANN AND E. PARDOUX

University of British Columbia and Université de Provence

It is shown that if a diffusion process, $\{X_t: 0 \le t \le 1\}$, on R^d satisfies

$$dX_t = b(t, X_t)\,dt + \sigma(t, X_t)\,dw_t$$

then the reversed process, $\{\overline{X}_t: 0 \le t \le 1\}$ where $\overline{X}_t = X_{1-t}$, is again a diffusion with drift $\bar{b}$ and diffusion coefficient $\bar{\sigma}$, provided some mild conditions on b, σ, and p_0, the density of the law of X_0, hold. Moreover $\bar{b}$ and $\bar{\sigma}$ are identified.

1. Brownian motion **trajectories** are **inherently irregular**, this **limits** how large an SDE solver's **step size** can be **without losing accuracy**

2. **Many sources of error** to control: **discretization** from finite step sizes + **injected noise**

$$x_{t_{i-1}} = x_{t_i} + \left[f(x_{t_i}, t_i) - g(t_i)^2 s_\theta(x_{t_i}, t_i)\right]\Delta t + g(t_i)\sqrt{\Delta t}\,\xi_i$$

$\sim \mathrm{N}(0, \mathbb{I})$

Euler–Maruyama Discretization

(...SDE solver inference)

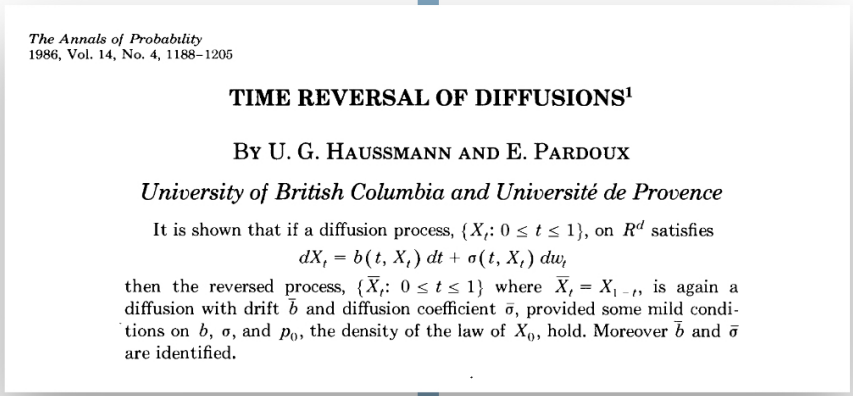


Flavor 2:
Score Matching, SDE

Backward Process

$$dx = \underbrace{f(x,t)dt}_{\texttt{drift}} + \underbrace{g(t)dW}_{\texttt{diffusion}}$$

PROBLEMS



The Annals of Probability
1986, Vol. 14, No. 4, 1188–1205

TIME REVERSAL OF DIFFUSIONS[1]

BY U. G. HAUSSMANN AND E. PARDOUX

University of British Columbia and Université de Provence

It is shown that if a diffusion process, $\{X_t: 0 \le t \le 1\}$, on R^d satisfies

$$dX_t = b(t, X_t)\,dt + \sigma(t, X_t)\,dw_t$$

then the reversed process, $\{\overline{X}_t: 0 \le t \le 1\}$ where $\overline{X}_t = X_{1-t}$, is again a diffusion with drift $\bar{b}$ and diffusion coefficient $\bar{\sigma}$, provided some mild conditions on b, σ, and p_0, the density of the law of X_0, hold. Moreover $\bar{b}$ and $\bar{\sigma}$ are identified.

1. Brownian motion **trajectories** are **inherently irregular**, this **limits** how large an SDE solver's **step size** can be **without losing accuracy**

2. **Many sources of error** to control: **discretization** from finite step sizes + **injected noise**

$$x_{t_{i-1}} = x_{t_i} + \left[f(x_{t_i}, t_i) - g(t_i)^2 s_\theta(x_{t_i}, t_i)\right]\Delta t + g(t_i)\sqrt{\Delta t}\,\xi_i$$

$$\texttt{Euler}-\texttt{Maruyama Discretization}$$

$$\sim \mathrm{N}(0, \mathbb{I})$$



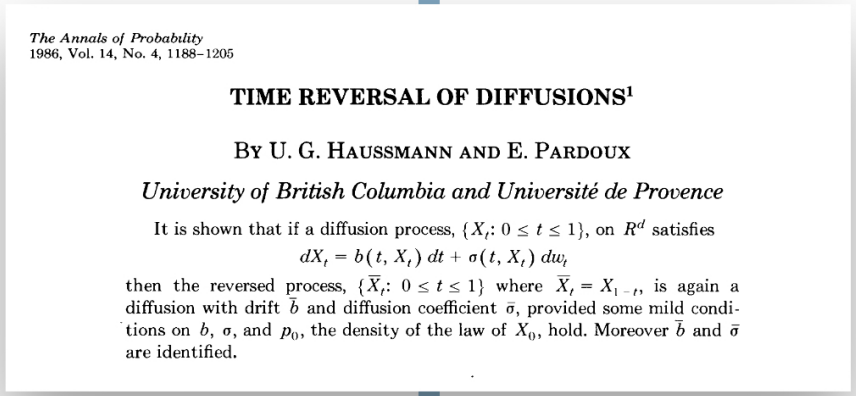
(...SDE solver convergence)

Flavor 2:
Score Matching, SDE

Backward Process

SOLUTION

$$dx = f(x,t)dt + g(t)dW$$



The Annals of Probability
1986, Vol. 14, No. 4, 1188–1205

TIME REVERSAL OF DIFFUSIONS[1]

BY U. G. HAUSSMANN AND E. PARDOUX

University of British Columbia and Université de Provence

It is shown that if a diffusion process, $\{X_t: 0 \le t \le 1\}$, on R^d satisfies

$$dX_t = b(t, X_t)\,dt + \sigma(t, X_t)\,dw_t$$

then the reversed process, $\{\overline{X}_t: 0 \le t \le 1\}$ where $\overline{X}_t = X_{1-t}$, is again a diffusion with drift $\bar{b}$ and diffusion coefficient $\bar{\sigma}$, provided some mild conditions on b, σ, and p_0, the density of the law of X_0, hold. Moreover $\bar{b}$ and $\bar{\sigma}$ are identified.

$$dx = \left[f(x,t) - g(t)^2 \nabla_x \log q_t(x) \right] dt + g(t) d\widetilde{W}$$

Stochastic Differential Equation

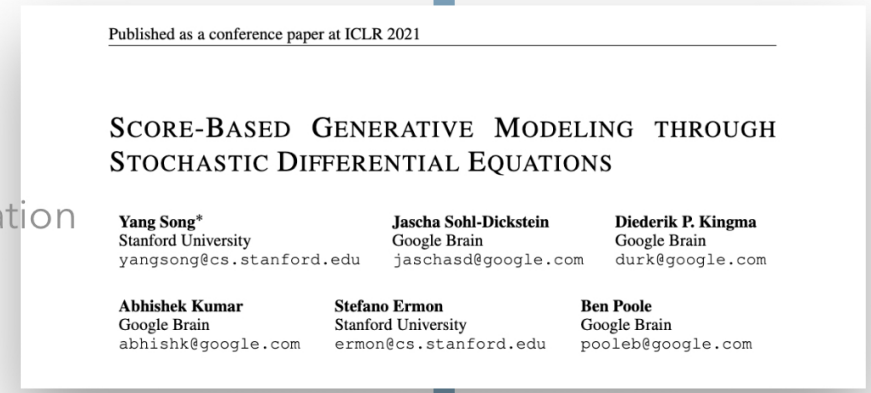
Flavor 2:
Score Matching, SDE

Backward Process

$$dx = f(x,t)dt + g(t)dW$$

SOLUTION

Fokker—Planck Equation



Published as a conference paper at ICLR 2021

SCORE-BASED GENERATIVE MODELING THROUGH STOCHASTIC DIFFERENTIAL EQUATIONS

Yang Song*
Stanford University
yangsong@cs.stanford.edu

Jascha Sohl-Dickstein
Google Brain
jaschasd@google.com

Diederik P. Kingma
Google Brain
durk@google.com

Abhishek Kumar
Google Brain
abhishk@google.com

Stefano Ermon
Stanford University
ermon@cs.stanford.edu

Ben Poole
Google Brain
pooleb@google.com

Probability Flow
Ordinary Differential Equation
(Neural-ODE)

$$dx = \left[f(x,t) - \frac{1}{2} g(t)^2 \nabla_x \log q_t(x) \right] dt + \varnothing$$

Ordinary Differential Equation

1. Reverse-SDE and PF-ODE generations **share** the same **time-marginal** q_t although now the **only** source of randomness is the initial condition x_T
2. Reverse-SDE and PF-ODE Individual **trajectories** are, instead, completely **different**
3. Being an **ODE**, it can now be integrated via **higher-order** and adaptive **solvers**: **only** finite step sizes contribute to error

Flavor 2:
Score Matching, SDE

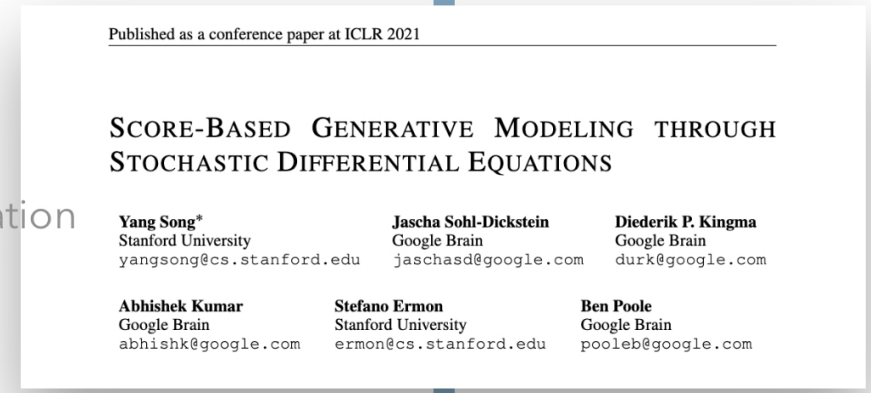
Backward Process

$$dx = f(x,t)dt + g(t)dW$$

SOLUTION

Fokker—Planck Equation

Probability Flow
Ordinary Differential Equation
(Neural-ODE)



Published as a conference paper at ICLR 2021

SCORE-BASED GENERATIVE MODELING THROUGH STOCHASTIC DIFFERENTIAL EQUATIONS

Yang Song*
Stanford University
yangsong@cs.stanford.edu

Jascha Sohl-Dickstein
Google Brain
jaschasd@google.com

Diederik P. Kingma
Google Brain
durk@google.com

Abhishek Kumar
Google Brain
abhishk@google.com

Stefano Ermon
Stanford University
ermon@cs.stanford.edu

Ben Poole
Google Brain
pooleb@google.com

	Reverse SDE	PF-ODE
Nature of process	Stochastic	Deterministic
Sampling diversity	Higher	Lower
Sampling quality	Higher	Lower
Sampling speed	Slower	Faster

$$dx = \left[f(x,t) - \frac{1}{2} g(t)^2 \nabla_x \log q_t(x) \right] dt + \varnothing$$

Ordinary Differential Equation



Flavor 3:
Flow Matching

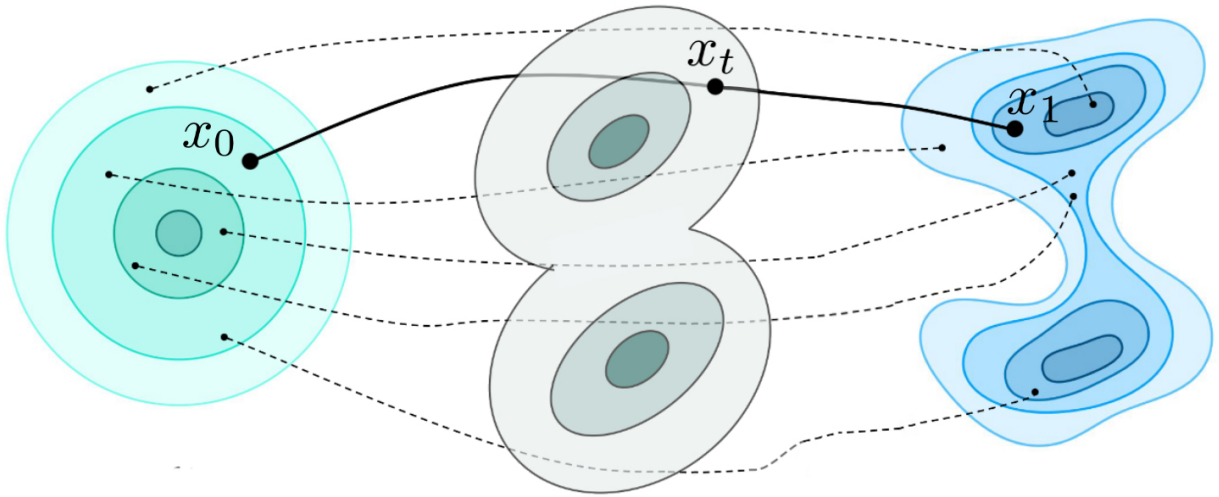
Let's start from where we stopped…

…that is, Fokker-Planck equation and the induced PF-ODE, shall we?

Flavor 3:
Flow Matching

SETUP

Flow $\psi_t(x_0)$: collection of trajectories starting from different initial point



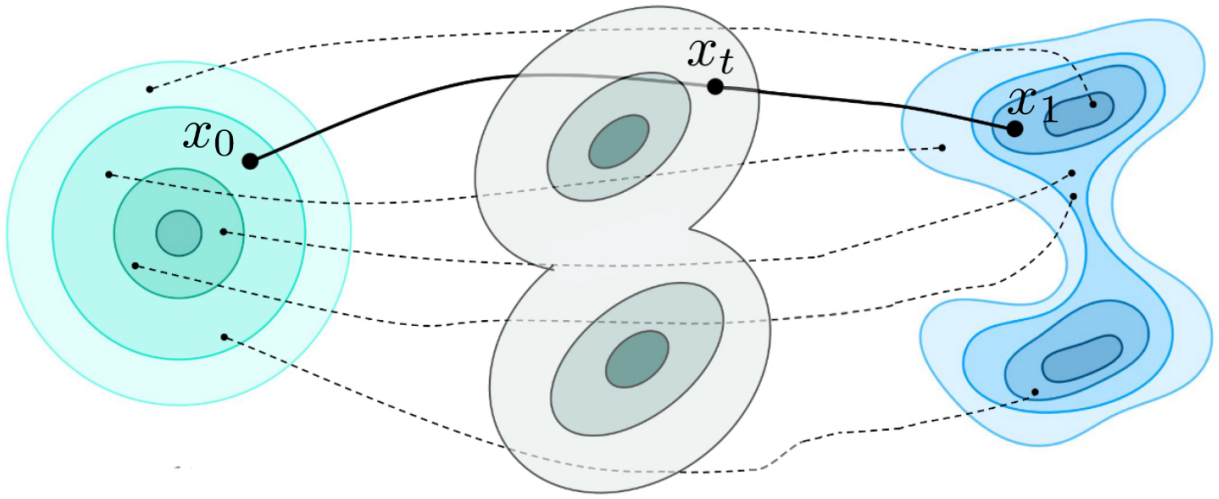
$q_0 = \mathrm{N}(0, \mathbb{I})$ $\quad$ q_t (distribution of x_t) $\quad$ $q_1 = p$

Transport Problem

Flavor 3:
Flow Matching

SETUP

Flow $\psi_t(x_0)$: collection of trajectories starting from different initial point

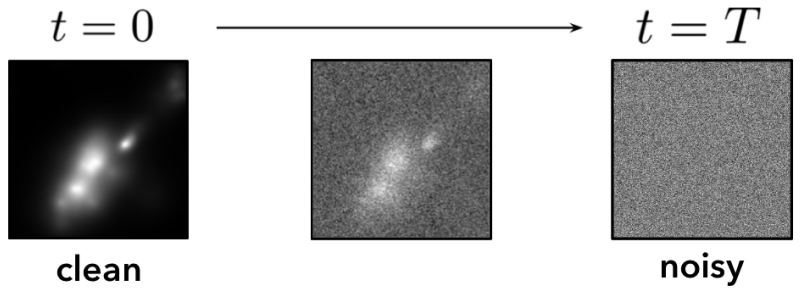


$q_0 = \mathrm{N}(0, \mathbb{I})$ $\quad$ q_t (distribution of x_t) $\quad$ $q_1 = p$

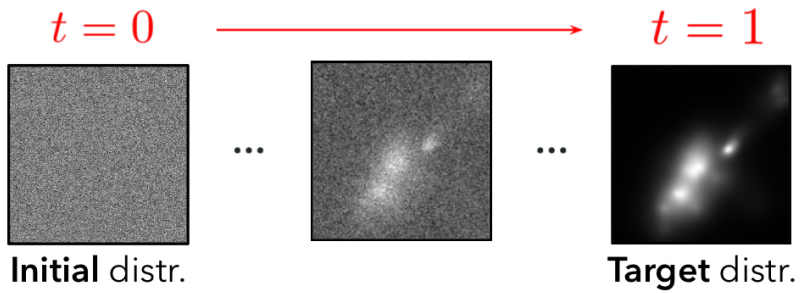
Transport Problem

CHANGE CONVENTION

Diffusion, Score Matching, SDE formulation



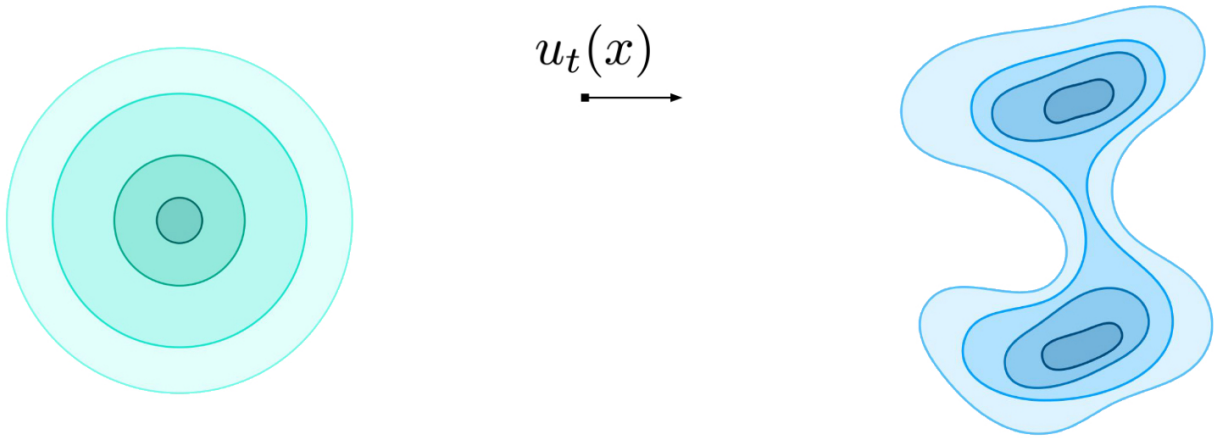
Flow Matching



Flavor 3:
Flow Matching

IDEA

Velocity or **Vector Field**: where to move (direction, speed) at any time and location



$q_0 = \mathrm{N}(0, \mathbb{I})$

$q_1 = p$

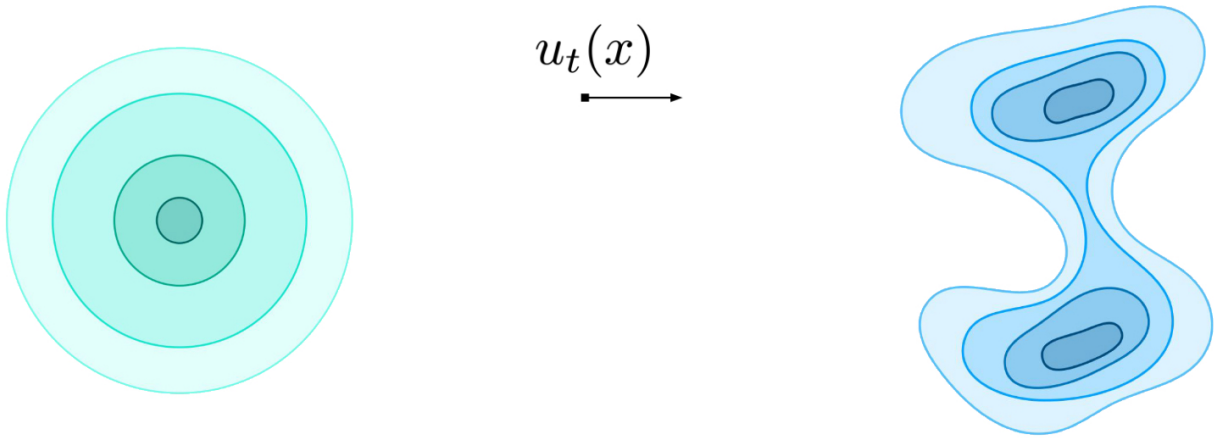
Velocity characterizes the Flow!

(…under smoothness conditions on the trajectories…)

Flavor 3:
Flow Matching

IDEA | DEJA-VU

Velocity or **Vector Field**: where to move (direction, speed) at any time and location



$q_0 = \mathrm{N}(0, \mathbb{I})$

$q_1 = p$

Velocity characterizes the Flow!

The **PF-ODE** already had a velocity field!

$$\underbrace{\frac{dx}{dt}}_{\text{rate of variation in time}} = \underbrace{\left[f(x,t) - \frac{1}{2} g(t)^2 \nabla_x \log q_t(x) \right]}_{\text{velocity}} \Leftrightarrow \frac{dx}{dt} = u_t(x)$$

Background **centering** inherited by the **drift** of the forward process

The **diffusivity** modulate the **strenght** of the score **correction** over time

Density-driven **correction** that pushes probability toward the data distribution

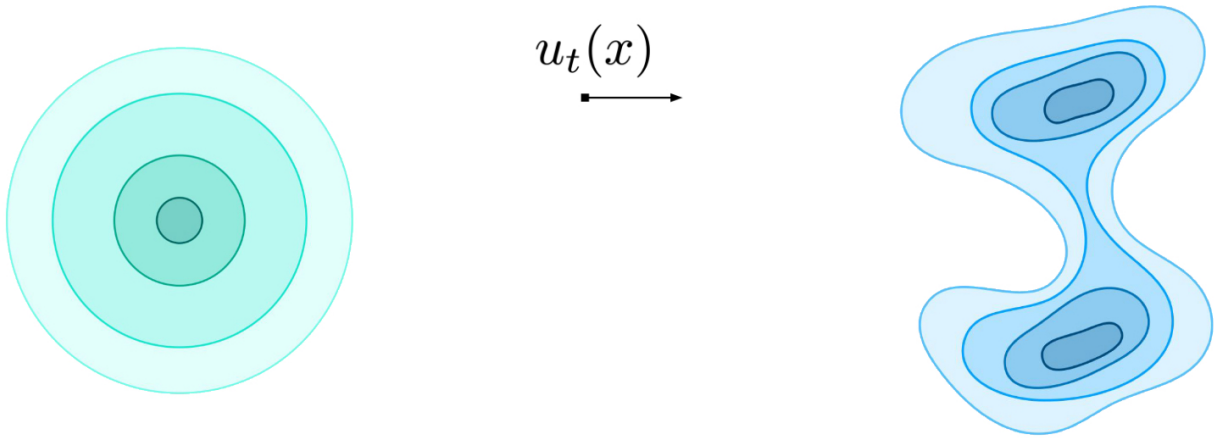
(…under smoothness conditions on the trajectories…)

Flavor 3:
Flow Matching

(...under smoothness conditions on the trajectories...)

IDEA ↩ DEJA-VU ↩

Velocity or **Vector Field**: where to move (direction, speed) at any time and location



Velocity characterizes <u>also</u> the probability path!

The **PF-ODE** already had a velocity field!

rate of variation in time = velocity

$$\frac{dx}{dt} = \left[f(x,t) - \frac{1}{2} g(t)^2 \nabla_x \log q_t(x) \right] \quad \Leftrightarrow \quad \frac{dx}{dt} = u_t(x)$$

Background **centering** inherited by the **drift** of the forward process

Density-driven **correction** that pushes probability toward the data distribution

The **diffusivity** modulate the **strenght** of the score **correction** over time

(after some algebra)

Continuity equation from Fokker-Planck
(a.k.a. conservation of mass)

Inflow - Outflow

Probability Flux (density x velocity)

$$\frac{\partial q_t(x)}{\partial t} + \nabla_x \cdot \left(q_t(x) u_t(x) \right) = 0$$

"Macroscopic" behaviour

"Microscopic" behaviour

$\mathrm{div}(h) = \nabla \cdot h = \sum_{j=1}^{p} \frac{\partial h}{\partial x_i}$

Flavor 3:
Flow Matching

IDEA | DEJA-VU

Velocity or **Vector Field**: where to move (direction, speed) at any time and location

The **PF-ODE** already had a velocity field!

New Goal

Learn directly the **velocity** (instead of the score) and then numerically solve the ODEs

$$\widehat{x}_1 = x_0 + \int_0^1 u_\theta(x_t, t)\, dt \quad \text{with} \quad \frac{dx_t}{dt} = u_\theta(x_t, t)$$

velocity

$$q_t(x) \quad \Leftrightarrow \quad \frac{dx}{dt} = u_t(x)$$

Density-driven **correction** that pushes probability toward the data distribution

odulate
e score
time

$$q_0 = \mathrm{N}(0, \mathbb{I})$$

$$q_1 = p$$

rom Fokker-Planck
ion of mass)

Inflow - Outflow

Probability Flux
(density x velocity)

$$\frac{\partial q_t(x)}{\partial t} + \nabla_x \cdot \left(q_t(x) u_t(x)\right) = 0$$

"Macroscopic" behaviour

"Microscopic" behaviour

$$\mathrm{div}(h) = \nabla \cdot h = \sum_{j=1}^{p} \frac{\partial h}{\partial x_i}$$

(...under smoothness conditions on the trajectories...)

Flavor 3:
Flow Matching

$$\mathcal{L}_{\text{FM}} = \mathbb{E}_{t,x}\left[\left\| u_\theta(x,t) - \underbrace{u(x,t)}_{\textbf{unknown!}} \right\|^2\right]$$

↬ FLOW MATCHING ↫

Published as a conference paper at ICLR 2023

FLOW MATCHING FOR GENERATIVE MODELING

Yaron Lipman[1,2] **Ricky T. Q. Chen**[1] **Heli Ben-Hamu**[2] **Maximilian Nickel**[1] **Matt Le**[1]
[1]Meta AI (FAIR) [2]Weizmann Institute of Science

ABSTRACT

We introduce a new paradigm for generative modeling built on Continuous Normalizing Flows (CNFs), allowing us to train CNFs at unprecedented scale. Specifically, we present the notion of Flow Matching (FM), a simulation-free approach for training CNFs based on regressing vector fields of fixed conditional probability paths. Flow Matching is compatible with a general family of Gaussian probability paths for transforming between noise and data samples—which subsumes existing diffusion paths as specific instances. Interestingly, we find

A. Derive velocity for a **simple target**
   ↬ **Dirac delta** on a single point

B. Construct velocity for **generic target**
   ↬ **Continuity equation**

C. Show loss function is **tractable**
   ↬ **Gradient** of FM and CFM are equal

D. Deduce loss function
   ↬ Very simple…even **too simple**!

↬ STRATEGY ↫

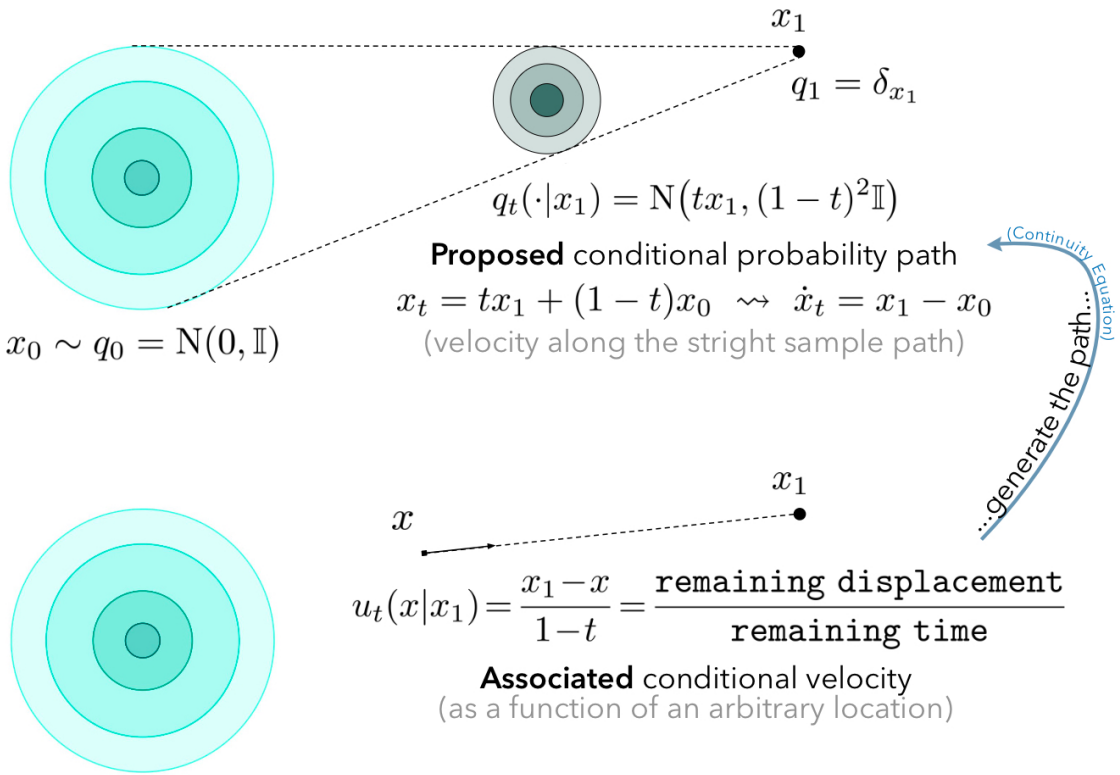
Flavor 3:
Flow Matching

$$\mathcal{L}_{\text{FM}} = \mathbb{E}_{t,x}\left[\left\| u_\theta(x,t) - u(x,t) \right\|^2\right]$$

unknown!

↬ FLOW MATCHING ↫

A — CONDITIONAL

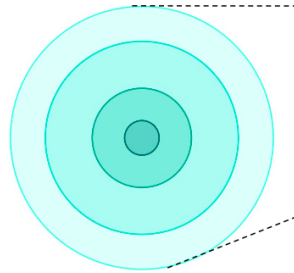


Flavor 3:
Flow Matching

$$\mathcal{L}_{\text{FM}} = \mathbb{E}_{t,x}\Big[\big\|u_\theta(x,t) - \underbrace{u(x,t)}_{\texttt{unknown!}}\big\|^2\Big]$$

↬ FLOW MATCHING ↫

A — CONDITIONAL



x_1

$q_1 = \delta_{x_1}$

$q_t(\cdot|x_1) = \mathrm{N}\big(tx_1, (1-t)^2\mathbb{I}\big)$

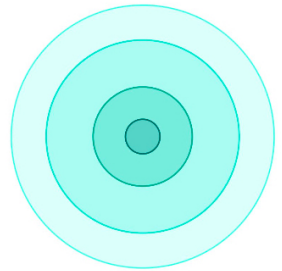
Proposed conditional probability path

$x_t = tx_1 + (1-t)x_0 \;\rightsquigarrow\; \dot{x}_t = x_1 - x_0$

(velocity along the stright sample path)

$x_0 \sim q_0 = \mathrm{N}(0, \mathbb{I})$

(Continuity Equation) ...generate the path...



x x_1

$u_t(x|x_1) = \dfrac{x_1 - x}{1-t} = \dfrac{\texttt{remaining displacement}}{\texttt{remaining time}}$

Associated conditional velocity

(as a function of an arbitrary location)

B — MARGINAL

Marginal probability path

$$q_t(x) = \int q_t(x \mid x_1)\, p(x_1)\, dx_1$$

(Def. of Marginal + Continuity Equation) ...generate the path...

Marginal velocity

$$u_t(x) = \int u_t(x \mid x_1)\, \pi(x_1|x)\, dx_1$$

$\texttt{posterior mean velocity}$

with $\pi(x_1|x) = \dfrac{q_t(x \mid x_1)\, p(x_1)}{q_t(x)}$

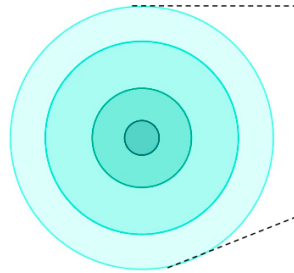
$\texttt{where to go knowing where you are}$

Flavor 3:
Flow Matching

$$\mathcal{L}_{\text{FM}} = \mathbb{E}_{t,x}\Big[\big\|u_\theta(x,t) - \underbrace{u(x,t)}_{\texttt{unknown!}}\big\|^2\Big]$$

FLOW MATCHING

A — CONDITIONAL



x_1

$q_1 = \delta_{x_1}$

$q_t(\cdot|x_1) = \mathrm{N}\big(tx_1, (1-t)^2\mathbb{I}\big)$

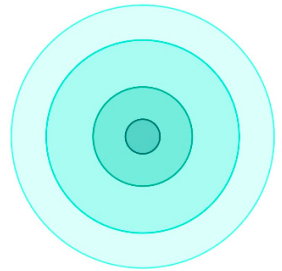
Proposed conditional probability path

$x_t = tx_1 + (1-t)x_0 \;\rightsquigarrow\; \dot{x}_t = x_1 - x_0$

(velocity along the stright sample path)

$x_0 \sim q_0 = \mathrm{N}(0, \mathbb{I})$

(Continuity Equation) ...generate the path...



x x_1

$$u_t(x|x_1) = \frac{x_1 - x}{1-t} = \frac{\texttt{remaining displacement}}{\texttt{remaining time}}$$

Associated conditional velocity

(as a function of an arbitrary location)

B — MARGINAL

Marginal probability path

$$q_t(x) = \int q_t(x \mid x_1)\, p(x_1)\, dx_1$$

(Def. of Marginal + Continuity Equation) ...generate the path...

Marginal velocity

$$u_t(x) = \int u_t(x \mid x_1)\, \pi(x_1|x)\, dx_1$$

posterior mean velocity

with $\pi(x_1|x) = \dfrac{q_t(x \mid x_1)\, p(x_1)}{q_t(x)}$

where to go knowing where you are

PNRR Missione 4 • Componente 2 Investimento 3.1
STILES - IR0000034 CUP C33C22000640006

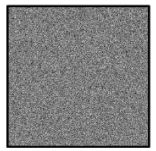
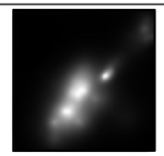
Flavor 3:
Flow Matching

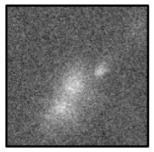
$$\mathcal{L}_{\text{CFM}} = \mathbb{E}_{t,x_1,x}\left[\left\| u_t^\theta(x) - \underbrace{u_t(x|x_1)}_{x_1 - x_0} \right\|^2\right]$$

CONDITIONAL FLOW MATCHING

TRAINING

1. Sample:

noise $x_0 \sim \mathrm{N}(0, \mathbb{I})$  **clean** image $x_1 \sim p$  **time step** $t \sim \mathcal{U}(0,1)$

noised image $x_t = (1-t)x_0 + tx_1$ 

2. Use x_t and t to **predict** $x_1 - x_0$ via $u_t^\theta(x_t)$

3. Compute **loss** $\mathcal{L} = \left\| u_t^\theta(x_t) - (x_1 - x_0) \right\|^2$ and **backpropagate**

INFERENCE

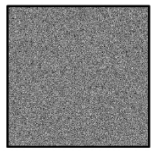
Flavor 3:
Flow Matching

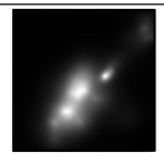
$$\mathcal{L}_{\text{CFM}} = \mathbb{E}_{t,x_1,x}\left[\left\| u_t^\theta(x) - \underbrace{u_t(x|x_1)}_{x_1 - x_0} \right\|^2\right]$$

CONDITIONAL FLOW MATCHING

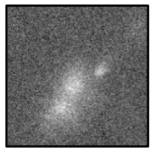
TRAINING

1. Sample:

noise $x_0 \sim \mathrm{N}(0, \mathbb{I})$ 

clean image $x_1 \sim p$ 

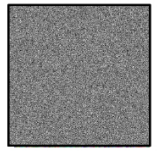
time step $t \sim \mathcal{U}(0,1)$

noised image $x_t = (1-t)x_0 + tx_1$ 

2. Use x_t and t to **predict** $x_1 - x_0$ via $u_t^\theta(x_t)$

3. Compute **loss** $\mathcal{L} = \left\| u_t^\theta(x_t) - (x_1 - x_0) \right\|^2$ and **backpropagate**

INFERENCE

1. Sample: **noise** $x_0 \sim \mathrm{N}(0, \mathbb{I})$ 

2. Use Euler to numerically solve the ODE:

$$x_{t_i} = x_{t_{i-1}} + u_{t_{i-1}}^\theta(x_{t_{i-1}})(t_i - t_{i-1})$$

3. Obtain **final** image x_1 

Flavor 3:
Flow Matching



Flow Straight and Fast:
Learning to Generate and Transfer Data with Rectified Flow

Xingchao Liu*
University of Texas at Austin
xcliu@utexas.edu

Chengyue Gong*
University of Texas at Austin
cygong@cs.utexas.edu

Qiang Liu
University of Texas at Austin
lqiang@cs.utexas.edu

arXiv:2209.03003v1 [cs.LG] 7 Sep 2022

Abstract

We present rectified flow, a surprisingly simple approach to learning (neural) ordinary differential equation (ODE) models to transport between two empirically observed distributions π_0 and π_1, providing a unified solution to generative modeling and domain transfer, among various tasks involving distribution transport. The idea of rectified flow is to learn the ODE to follow straight paths connecting the points drawn from π_0 and π_1 as much as possible. This is achieved by straightforward nonlinear least squares optimization, which can be easily scaled to large models without introducing extra parameters beyond standard supervised learning. The straight paths are the optimal transport (in L_2 sense) between π_0 and π_1, and hence rectified flow also reduces the transport cost. In addition, rectified flow straightens any paths through successive reflow procedures, which eventually yield straight paths that are both optimal and deterministic couplings, and can be simulated exactly with a single Euler step. Moreover, rectified flow can be applied to learn deterministic coupling with provably non-increasing convex transport costs. In addition, rectified flow can be simulated accurately with coarse time discretization in practice. Empirically, we show that rectified flow performs superbly on image generation, image-to-image translation, and domain adaptation. In particular, on image generation and translation, our method yields nearly straight flows that give high quality results even with *a single Euler step*.

1 Introduction

Compared with supervised learning, the shared

lack of *paired* input/output data with which st

gist of most unsupervised methods is to find

points from two distributions. For exam

(GAN) and variational autoencoders (VA

lowing a simple elementary (Gaussian)

Representation learning rests on the

Published as a conference paper at ICLR 2025

GENERATOR MATCHING: GENERATIVE MODELING WITH ARBITRARY MARKOV PROCESSES

Peter Holderrieth[1,†], **Marton Havasi**[2], **Jason Yim**[1], **Neta Shaul**[2,3], **Itai Gat**[2],
Tommi Jaakkola[1], **Brian Karrer**[2], **Ricky T. Q. Chen**[2], **Yaron Lipman**[2]
[1]MIT CSAIL, [2]FAIR, Meta, [3]Weizmann Institute of Science
[†]Work done during an internship at FAIR, Meta.

ABSTRACT

We introduce *Generator Matching*, a modality-agnostic framework for generative modeling using arbitrary Markov processes. Generators characterize the infinitesimal evolution of a Markov process, which we leverage for generative modeling in a similar vein to flow matching: we construct conditional generators which generate single data points, then learn to approximate the marginal generator which generates the full data distribution. We show that Generator Matching unifies various generative modeling methods, including diffusion models, flow matching and discrete diffusion models. Furthermore, it expands the design space to new and unexplored Markov processes such as jump processes. Finally, Generator Matching enables the construction of superpositions of Markov generative models and enables the construction of multimodal models in a rigorous manner. We empirically validate our method on image and multimodal generation, e.g. showing that superposition with a jump process improves performance.

1 INTRODUCTION

Early deep generative models—like VAEs (Kingma, 2013) and GANs (Goodfellow et al., 2014) generated samples in a single forward pass. With denoising diffusion models (DDMs) (Song et al., 2020; Ho et al., 2020), a paradigm shift happened were *step-wise* updates are used to transform noise into data. Similarly, scalable training of continuous normalizing flows (CNFs; Chen et al. 2018) via flow matching (Lipman et al., 2022; Liu et al., 2022; Albergo et al., 2023) allowed for high-quality and fast generative modeling by simulating an ODE. Since then, similar constructions based on diffusion and flows have also been extended to other modalities such as discrete data (Campbell et al., 2022; Gat et al., 2024) or data on manifolds (Chen & Lipman, 2024) leading to a variety of new generative models (Huang et al., 2022; Chen & Lipman, 2024).

The single common property of the aforementioned generative models is that they all construct a time-dependent vector field u_t starting with a sample $X_0 \sim p_{\text{simple}}$ from an easy-to-sample distribution p_{simple} and then following the ODE to obtain samples X_{t+h} of the next time step dependent on the current sample X_t and the vector field u_t. This also means that they are all *Markov processes*, i.e. the next step X_{t+h} only depends on the current step X_t and not the previous steps X_s for $s < t$. Moreover, they all rely on that Markov process having the right infinitesimal evolution, i.e. the vector field u_t has to be chosen such that the Markov process follows the desired distribution path p_t through time. Constructing such vector fields has been the core challenge of the field and has been approached differently for each of the mentioned models.

Finanziato dall'Unione europea
NextGenerationEU

Ministero dell'Università e della Ricerca

Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA

INAF
ISTITUTO NAZIONALE DI ASTROFISICA



PNRR
Missione 4 • Componente 2
Investimento 3.1

STILES – IR0000034
CUP C33C22000640006

Where were we?

- ...and beyond!

  Pretty much the same Langevin-based paradigm can be adapted to deal with a variety of **inverse problems** in order to recover a signal from **corrupted measurements**:

$$\boldsymbol{y} = \mathsf{H}\boldsymbol{x} + \boldsymbol{\epsilon}$$

  - De-blurring
  - In-Painting
  - De-Mosaicing
  - Super-Resolution

- Power is nothing without control
  - Often we'd to ensure the generated data not only looks realistic but also adheres to desired **connectivity** constraints.
  - With **topological steering** we guide the denoising process to produce outputs with specific structural properties (e.g. connected components, holes, or specific branching structures)
  - Typical tools **persistent homology** or **topological loss** functions, although they are quite **computationally intensive**

- **...and beyond!**

  Pretty much the same Langevin-based paradigm can be adapted to deal with a variety of **inverse problems** in order to recover a signal from **corrupted measurements**:

$$\boldsymbol{y} = \mathsf{H}\boldsymbol{x} + \boldsymbol{\epsilon}$$

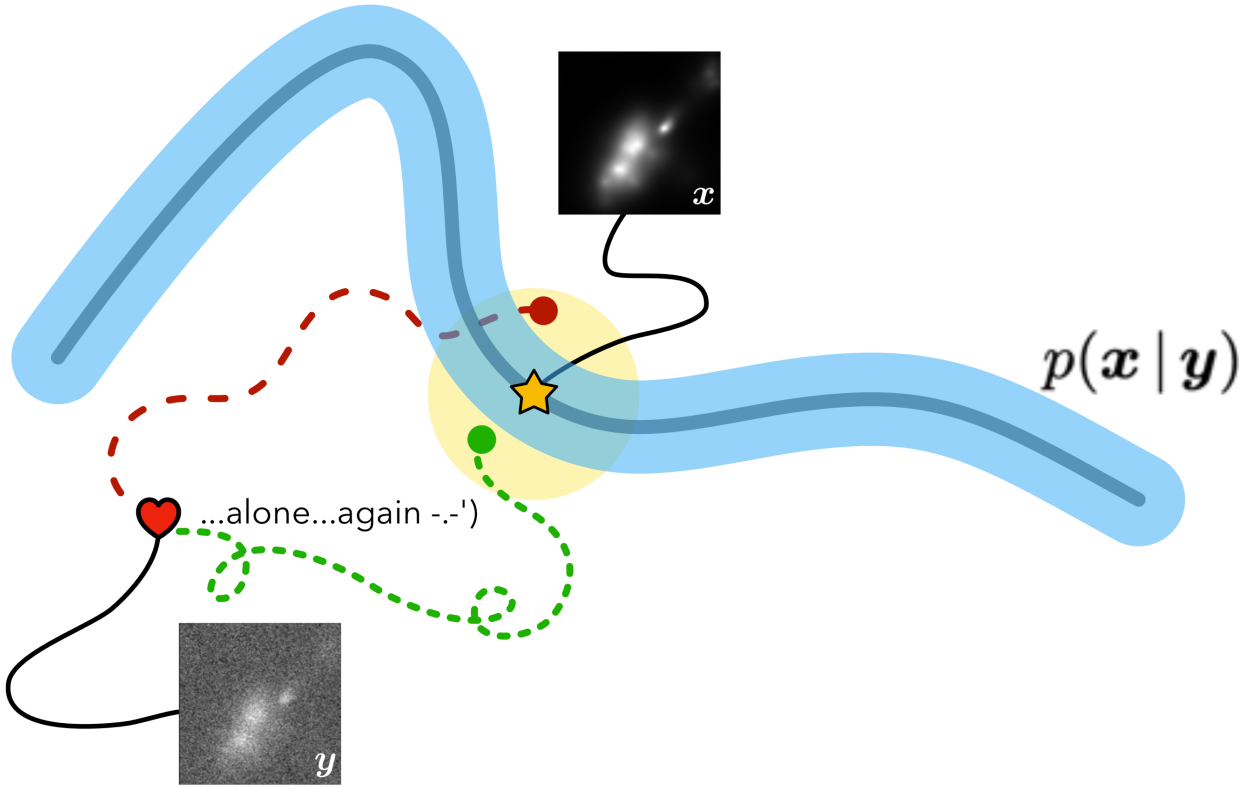
  - De-blurring
  - In-Painting
  - De-Mosaicing
  - Super-Resolution

- **Power is nothing without control**
  - Often we'd to ensure the generated data not only looks realistic but also adheres to desired **connectivity** constraints.
  - With **topological steering** we guide the denoising process to produce outputs with specific structural properties (e.g. connected components, holes, or specific branching structures)
  - Typical tools **persistent homology** or **topological loss** functions, although they are quite **computationally intensive**

- Denoising by Sampling:
  The Annealed Langevin algorithm can be twisted to design a **Stochastic Image Denoiser** that recover a sequence of gradually less noisy images by targeting the posterior distribution

$$p(\boldsymbol{x} \mid \boldsymbol{y})$$

  Solution is **not** unique.



Finanziato dall'Unione
Italiadomani
PIANO NAZIONALE
INAF
ISTITUTO NAZIONALE DI ASTROFISICA

Recent Surveys

Briding Diffusion Posterior Sampling and Monte Carlo methods: a survey

Y. Janati[1], A. Durmus[2], J. Olsson[3] and E. Moulines[4]
[1]Institute of Foundation Models Paris, MBZUAI
[2]KTH Royal Institute of Technology,
[3]Ecole polytechnique,
[4]MBZUAI

4114v1 [cs.LG] 15 Oct 2025

PHILOSOPHICAL TRANSACTIONS A
royalsocietypublishing.org/journal/rsta
Review
Cite this article: Milanfar P, Delbracio M. 2025 Denoising: a powerful building block for imaging, inverse and machine f 383: 20240326. .2024.0326

Denoising: a powerful building block for imaging, inverse problems and machine learning

Peyman Milanfar and Mauricio Delbracio
Google Inc, Mountain View, CA, USA
PM, 0000-0003-1455-7662

Denoising, the process of reducing random fluctuations in a signal to emphasize essential patterns, has been a fundamental problem of interest since the dawn of modern scientific inquiry. Recent denoising techniques, particularly in imaging, have achieved remarkable success, nearing theoretical limits by some measures. Yet, despite tens of thousands of research papers, the wide-ranging applications of denoising beyond noise removal have not been fully recognized. This is partly due to the vast and diverse literature, making a clear overview challenging. This perspective aims to address this

PHILOSOPHICAL TRANSACTIONS A
royalsocietypublishing.org/journal/rsta
Introduction
Cite this article: Oliviero-Durmus A, Janati Y, Moulines E, Pereyra M, Reich S. 2025 Generative modelling meets Bayesian inference: a new paradigm for inverse . R. Soc. A 383: 20240334.

Generative modelling meets Bayesian inference: a new paradigm for inverse problems

Alain Oliviero-Durmus[1], Yazid Janati[2], Eric Moulines[2], Marcelo Pereyra[3] and Sebastian Reich[4]
[1]Centre de Mathématiques Appliquées, Ecole Polytechnique, Palaiseau, Île-de-France, France
[2]Ecole Polytechnique, Palaiseau, Île-de-France
[3]Heriot–Watt University, Edinburgh,
[4]University of Potsdam, Potsdam, Ge
YJ, 0000-0002-7271-4104; EM, SR, 0000-0002-5336-8904

PHILOSOPHICAL TRANSACTIONS A
royalsocietypublishing.org/journal/rsta
Review
Cite this article: Luo Z, Gustafsson F, Zhao Z, Sjölund J, Schön T. 2025 Taming diffusion models for image restoration: a review. Phil Trans. R. Soc. A 383: 20240

Taming diffusion models for image restoration: a review

Ziwei Luo[1], Fredrik Gustafsson[2], Zheng Zhao[1,3], Jens Sjölund[1] and Thomas Schön[1]
[1]Department of Information Technology, Uppsala University, Uppsala, Sweden
[2]Department of Medical Epidemiology and Biostatistics, Karolinska Stockholm, Sweden
[3]Department of Computer and Inform Linköping, Sweden

A Survey on Diffusion Models for Inverse Problems

Giannis Daras[1], Hyungjin Chung[2], Chieh-Hsin Lai[3], Yuki Mitsufuji[3], Jong Chul Ye[2], Peyman Milanfar[4], Alexandros G. Dimakis[1], Mauricio Delbracio[4]
[1]UT Austin [2]KAIST [3]Sony AI [4]Google

Abstract

Diffusion models have become increasingly popular for generative modeling due to their ability to generate high-quality samples. This has unlocked exciting new possibilities for solving inverse problems, especially in image restoration and reconstruction, by treating diffusion models as unsupervised priors. This survey provides a comprehensive overview of methods that utilize pre-trained diffusion models without requiring further training. We introduce

STILES – IR0000034
CUP C33C22000640006

Finanziato dall'Unione europea
Italiadomani
PIANO NAZIONALE RIPRESA E RESILIENZA
INAF
ISTITUTO NAZIONALE DI ASTROFISICA

DDRM

Denoising Diffusion Restoration Models

Bahjat Kawar
Department of Computer Science
Technion, Haifa, Israel
bahjat.kawar@cs.technion.ac.il

Michael Elad
Department of Computer Science
Technion, Haifa, Israel
elad@cs.technion.ac.il

Jiaming Song
NVIDIA
Santa Clara, California, USA
jiamings@nvidia.com

DPS

Diffusion Posterior Sampling for General Noisy Inverse Problems

Hyungjin Chung[*1,2], **Jeongsol Kim**[*1], **Michael T. Mccann**[2], **Marc L. Klasky**[2] & **Jong Chul Ye**[1]
[1]KAIST, [2] Los Alamos National Laboratory
{hj.chung, jeongsol, jong.ye}@kaist.ac.kr, {mccann, mklasky}@lanl.gov

ABSTRACT

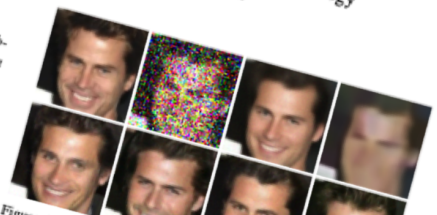
...udied as powerful generative inverse problem
...ons and the ease of combining exist-
...ing simple linear inverse
...the complexi...

Stochastic Image Denoising by Sampling from the Posterior Distribution

Bahjat Kawar Gregory Vaksman Michael Elad
Computer Science Department, The Technion - Israel Institute of Technology
{bahjat.kawar,grishav,elad}@cs.technion.ac.il

Abstract

Image denoising is a well-known and well studied problem, commonly targeting a minimization of the mean squared error (MSE) between the outcome and the original image. Unfortunately, especially for severe noise levels, such Minimum MSE (MMSE) solutions may lead to blurry output images. In this work we propose a novel stochastic denoising approach that produces viable and high perceptual quality results, while maintaining a small MSE. Our method employs Langevin dynamics that relies on a repeated application of any given MMSE denoiser, obtaining the reconstructed image by effectively sampling from the posterior distribution. Due to its stochastic nature, the algorithm can produce a variety of high perceptual quality images for a given noisy input.



[eess.IV] 31 Aug 2021

MCGdiff

...ished as a conference paper at ICLR 2024

Monte Carlo Guided Denoising Diffusion Models for Bayesian Linear Inverse Problems

Gabriel Cardoso*
Ecole polytechnique
IHU Lirye

Yazid Janati*
Ecole polytechnique

Sylvain Le Corff
Sorbonne Université

Eric Moulines
Ecole polytechnique

ABSTRACT

...ed linear inverse problems arise frequently in various applications, from computational
...dical imaging. A recent line of research exploits Bayesian inference with
...ill-posedness of such problems. Amongst such priors, score-
...been successfully applied to several different
...structure of the prior defined by
...As the noise level

InvFusion: Bridging Supervised and Zero-shot Diffusion for Inverse Problems

Noam Elata*
Technion
Haifa, Israel
noamelata@campus.technion.ac.il

Hyungjin Chung*
EverEx
Seoul, South Korea
hj.chung@everex.ac.kr

...ul Ye
...ST
...th Korea
...st.ac.kr

Tomer Michaeli
Technion
Haifa, Israel
tomer.m@ee.technion.ac.il

Michael Elad
Technion
Haifa, Israel
elad@cs.technion.ac.il

Abstract

Models have demonstrated remarkable capabilities in handling inverse
offering high-quality posterior-sampling-based solutions. Despite signif-
ces, a fundamental trade-off persists regarding the way the conditioned
employed: Zero-shot approaches can accommodate any linear degra-
r on approximations that reduce accuracy. In contrast, traini...
del the posterior correctly, but cannot adapt to the d...
we introduce InvFusion, the first training-...
apler. InvFusion combines the bes...
supervised approaches and ...
hrough a novel ar...
perator di...

Focus!

STILES – IR0000034
CUP C33C22000640006

DPS DDRM MCGdiff

Shared Perspective

Take an unconditional diffusion prior + "Steer" its reverse dynamics based on data

PNRR Missione 4 • Componente 2 Investimento 3.1
STILES – IR0000034 CUP C33C22000640006

DPS DDRM MCGdiff

Shared Perspective

Take an unconditional diffusion prior + "Steer" its reverse dynamics based on data

Example: score based 1/2

Suppose

$$y = Hx_0 + \epsilon, \qquad \epsilon \sim \mathrm{N}(0, \sigma_y^2 \mathbb{I}),$$

and our trained diffusion model supplies the unconditional time-dependent score

$$s_\theta(x_t, t) \approx \nabla_{x_t} \log p_t(x_t).$$

Without measurements, reverse diffusion uses this prior score to generate

$$x_T \sim \mathrm{N}(0, \mathbb{I}) \quad \longrightarrow \quad x_0 \sim p(x_0).$$

DPS DDRM MCGdiff

Shared Perspective

Take an unconditional diffusion prior + "Steer" its reverse dynamics based on data

Example: score based 1/2

Suppose

$$y = Hx_0 + \epsilon, \qquad \epsilon \sim \mathrm{N}(0, \sigma_y^2 \mathbb{I}),$$

and our trained diffusion model supplies the unconditional time-dependent score

$$s_\theta(x_t, t) \approx \nabla_{x_t} \log p_t(x_t).$$

Without measurements, reverse diffusion uses this prior score to generate

$$x_T \sim \mathrm{N}(0, \mathbb{I}) \longrightarrow x_0 \sim p(x_0).$$

Example: score based 2/2

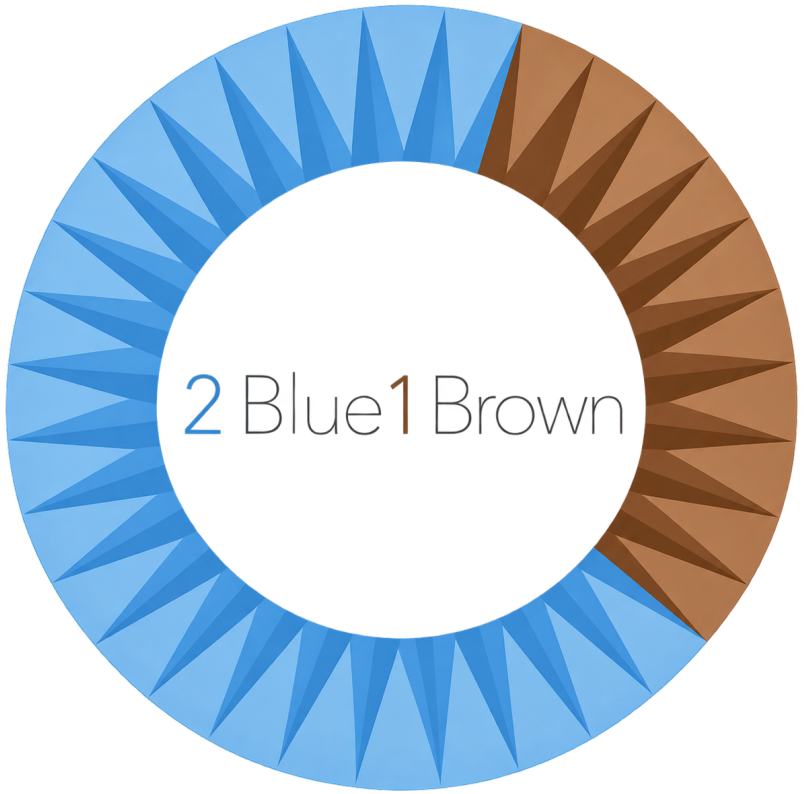
For posterior sampling we instead want

$$p_t(x_t \mid y).$$

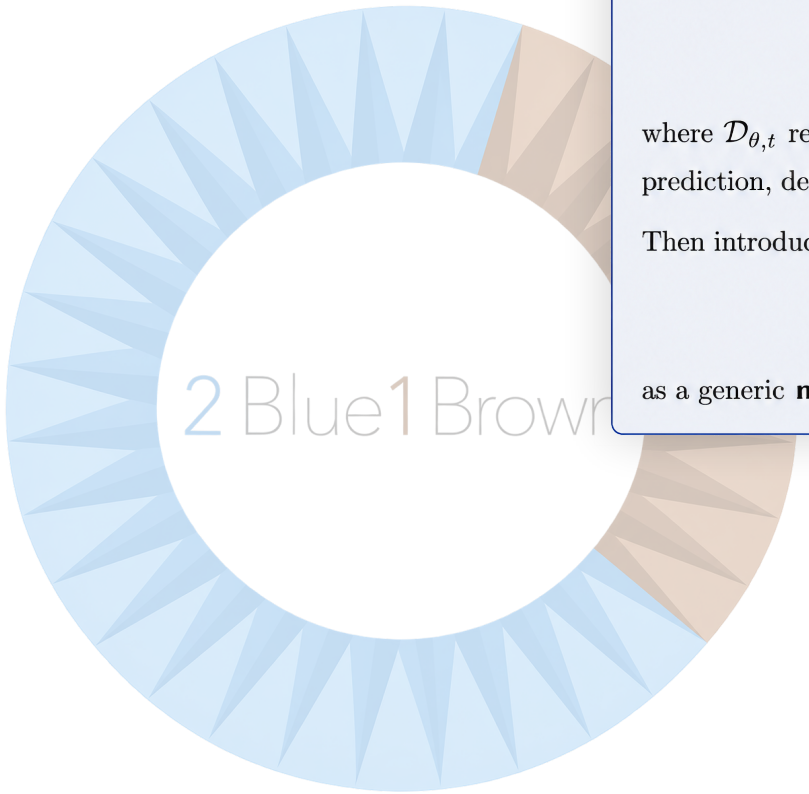
Bayes gives the fundamental decomposition

$$\nabla_{x_t} \log p_t(x_t \mid y) = \underbrace{\nabla_{x_t} \log p_t(x_t)}_{\text{prior score}} + \underbrace{\nabla_{x_t} \log p_t(y \mid x_t)}_{\text{measurement guidance}}$$

…the conditional score is, in general, **intractable**



Source: https://www.3blue1brown.com/



Example: reverse process & measurement conditioning 1/2

Write the unconditional reverse process schematically as

$$\boxed{\mathcal{D}_{\theta,t}(x_t)}$$

where $\mathcal{D}_{\theta,t}$ represents whatever the diffusion parameterization uses—score, noise prediction, denoiser, DDIM transition, reverse SDE drift, etc.

Then introduce

$$\boxed{\mathcal{G}_t(y, H, \sigma_y; x_t)}$$

as a generic **measurement-conditioning mechanism**.

Example: reverse process & measurement conditioning 2/2

Posterior sampling becomes schematically

$$\boxed{x_{t-1} = \text{Update}\left(\underbrace{\mathcal{D}_{\theta,t}(x_t)}_{\text{learned prior dynamics}}, \underbrace{\mathcal{G}_t(y, H, \sigma_y; x_t)}_{\text{measurement guide}}\right).}$$

Source: https://www.3blue1brown.com/

2 Blue 1 Brown

Example: reverse process & measurement conditioning 1/2

Write the unconditional reverse process schematically as

$$\boxed{\mathcal{D}_{\theta,t}(x_t)}$$

where $\mathcal{D}_{\theta,t}$ represents whatever the diffusion parameterization uses—score, noise prediction, denoiser, DDIM transition, reverse SDE drift, etc.

Then introduce

$$\boxed{\mathcal{G}_t(y, H, \sigma_y; x_t)}$$

as a generic **measurement-conditioning mechanism**.

Example: reverse process & measurement conditioning 2/2

Posterior sampling becomes schematically

$$\boxed{x_{t-1} = \text{Update}\left(\underbrace{\mathcal{D}_{\theta,t}(x_t)}_{\text{learned prior dynamics}}, \underbrace{\mathcal{G}_t(y, H, \sigma_y; x_t)}_{\text{measurement guide}}\right).}$$

Example: three algorithms via different $\mathcal{G}_t$

Then the three algorithms correspond to different realizations of $\mathcal{G}_t$:

- $\mathcal{G}_t^{\text{DPS}}$ = likelihood/discrepancy gradient through $\hat{x}_0(x_t)$,
- $\mathcal{G}_t^{\text{DDRM}}$ = SVD/operator-aware conditional transition,
- $\mathcal{G}_t^{\text{MCG}}$ = Monte-Carlo/particle conditioning correction.

Source: https://www.3blue1brown.com/

Finanziato dall'Unione europea
NextGenerationEU

Ministero dell'Università e della Ricerca

Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA

INAF
ISTITUTO NAZIONALE DI ASTROFISICA

OPEN UNIVERSE 2024



Data Description

- Simulated sky data for **Nancy Grace Roman Space Telescope** Wide Field Instrument (WFI), specifically the Roman Time Domain Survey (TDS).
- Images are **grayscale**, single-band astronomical FITS images. The selected band is R062 / F062, the relatively blue Roman wide-field imaging band.
- Restricting to one, well populated band avoids mixing different wavelength-dependent image distributions and PSFs while training the **diffusion prior**.
- Images contain simulated astronomical scenes including: stars / point sources; galaxies and extended objects; blends and crowded regions; background regions.

Pierpaolo Brutti | Sapienza

PNRR
Missione 4 • Componente 2
Investimento 3.1

STILES – IR0000034
CUP C33C22000640006

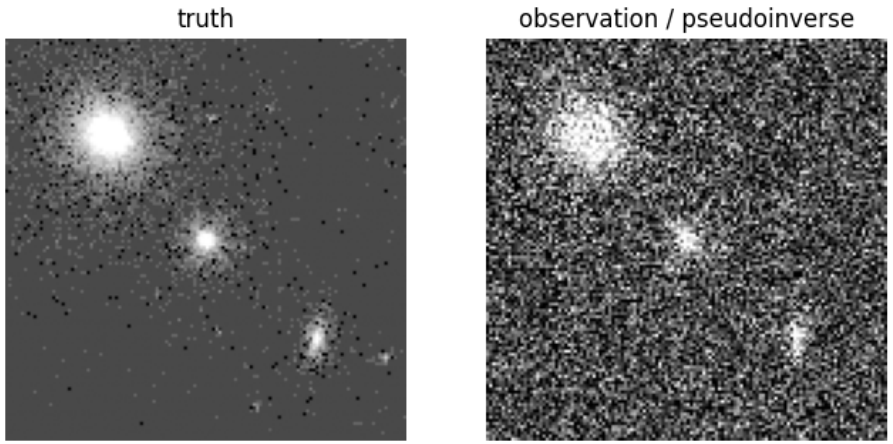
Finanziato dall'Unione europea
NextGenerationEU

Ministero dell'Università e della Ricerca

Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA

INAF
ISTITUTO NAZIONALE DI ASTROFISICA

OPEN UNIVERSE



	Training	Validation	Test
Pointings	12	3	4
FITS images	36	9	12
Patch size	128x128	128x128	128x128
Patch selection	Dynamic	Fixed	Fixed
Source/background	~70/30%	~75/25%	Source + background
Number of patches	**4608/epoch**	**~1152**	**1664 available**
Primary role	DDPM prior training	model tuning	final evaluation

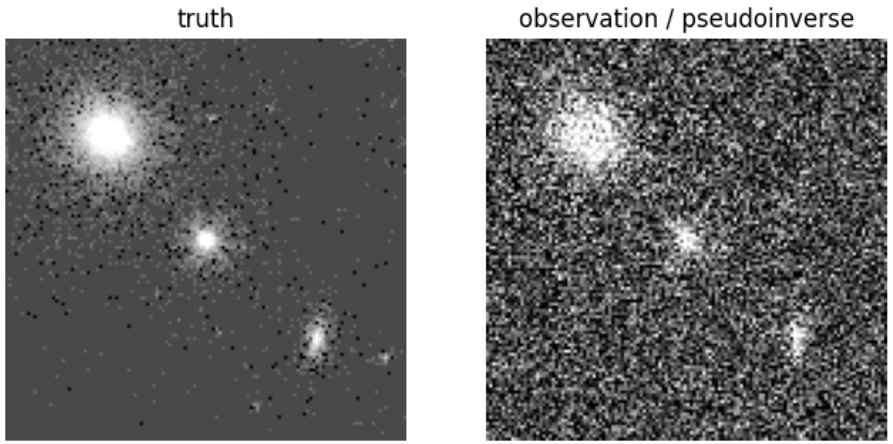
Finanziato dall'Unione europea
NextGenerationEU

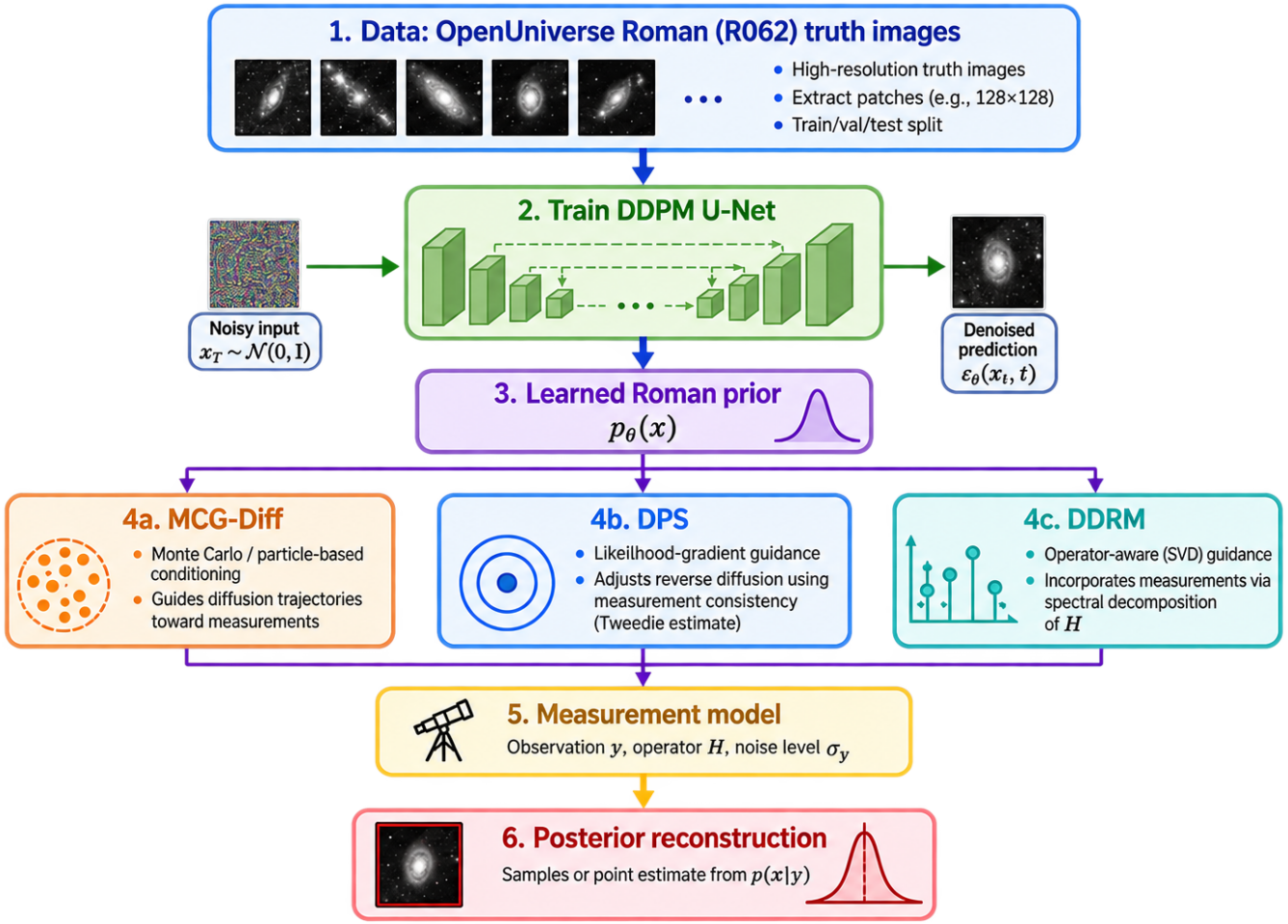
Ministero dell'Università e della Ricerca

Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA

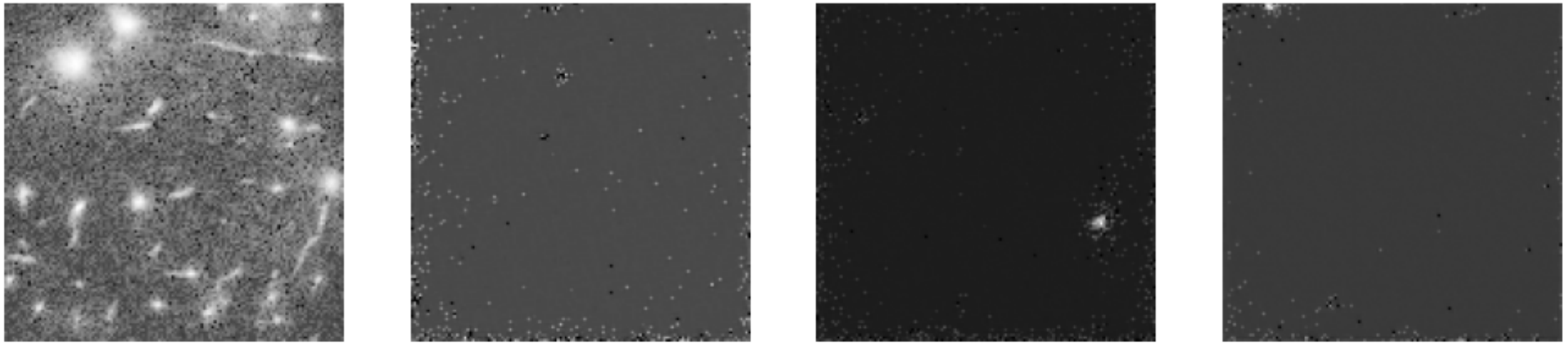
INAF
ISTITUTO NAZIONALE DI ASTROFISICA

OPEN UNIVERSE

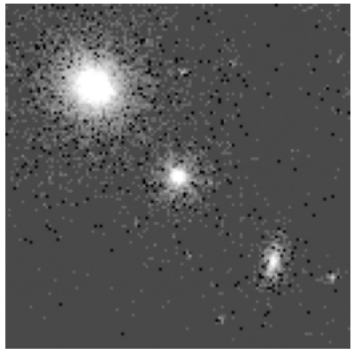




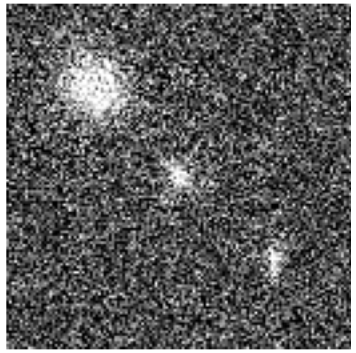
Unconditional DDPM samples: OpenUniverse RomanTDS R062



Truth

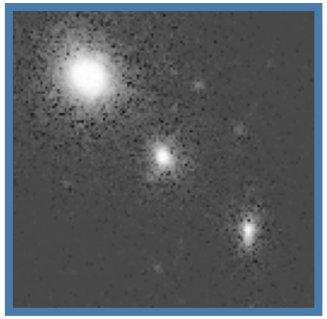


Observation / pinv

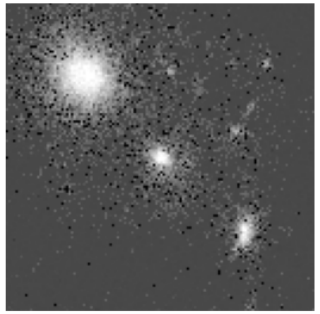


4 out of 8 draws

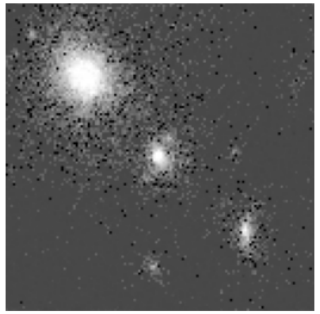
MCG-Diff mean



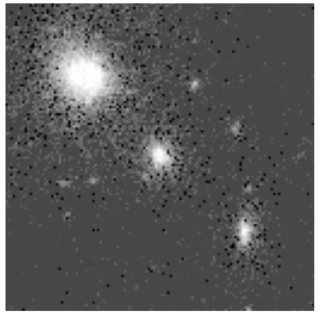
MCG-Diff draw 1



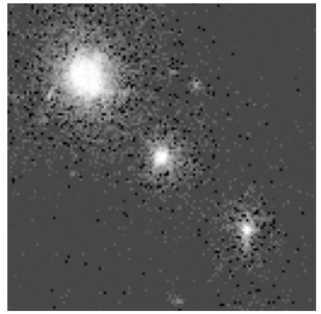
MCG-Diff draw 2



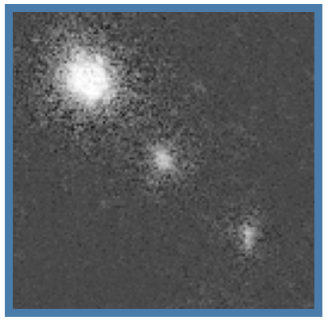
MCG-Diff draw 3



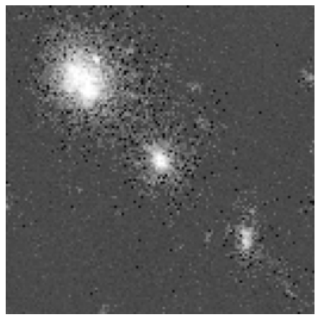
MCG-Diff draw 4



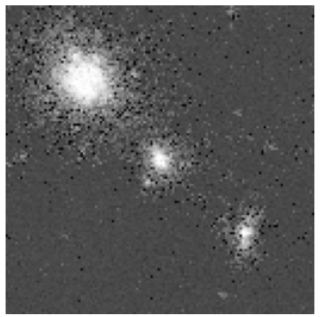
DPS mean



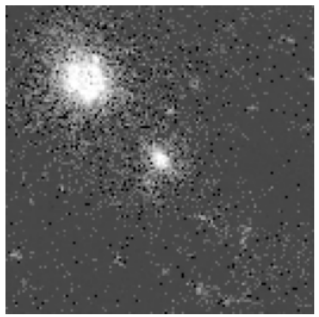
DPS draw 1



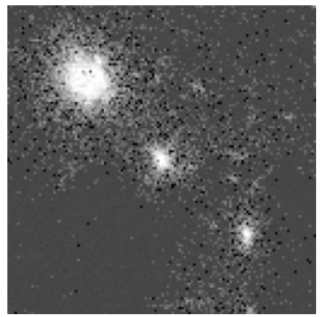
DPS draw 2



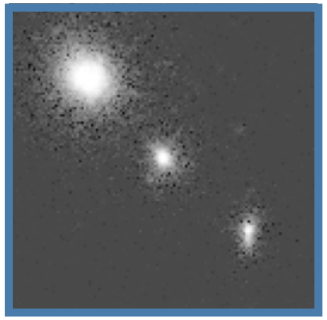
DPS draw 3



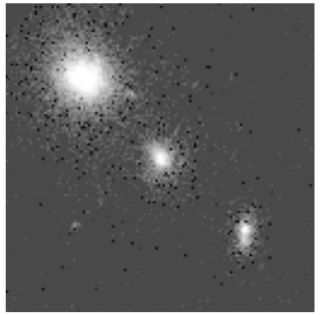
DPS draw 4



DDRM mean



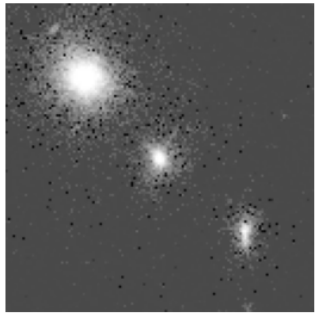
DDRM draw 1



DDRM draw 2



DDRM draw 3



DDRM draw 4



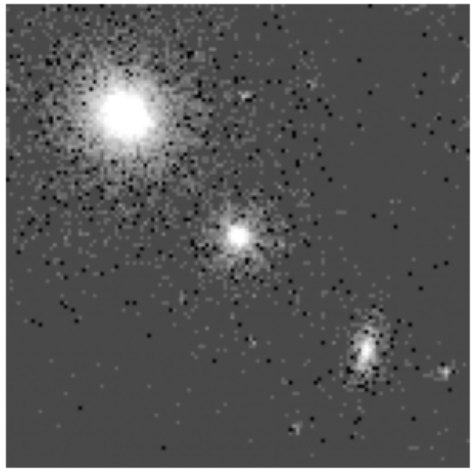
Finanziato dall'Unione europea
NextGenerationEU

Ministero dell'Università e della Ricerca

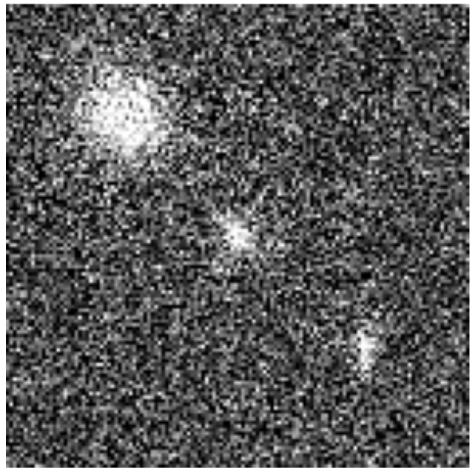
Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA

INAF
ISTITUTO NAZIONALE DI ASTROFISICA

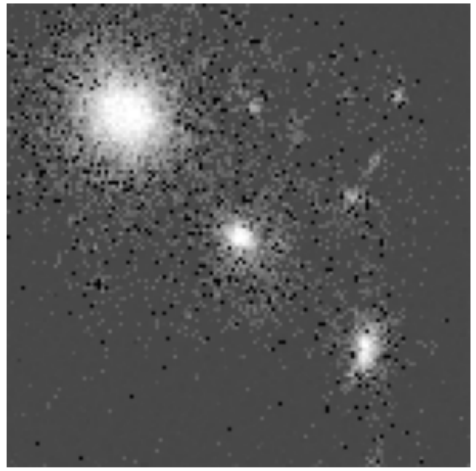
truth



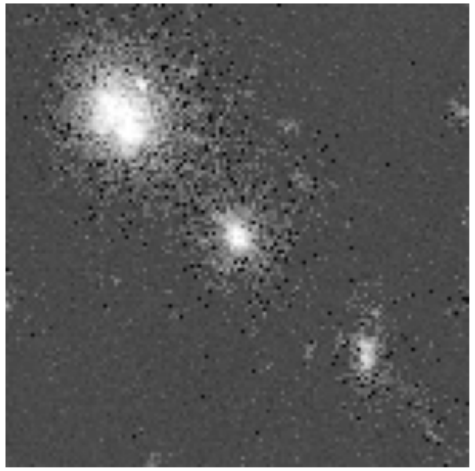
observation / pinv



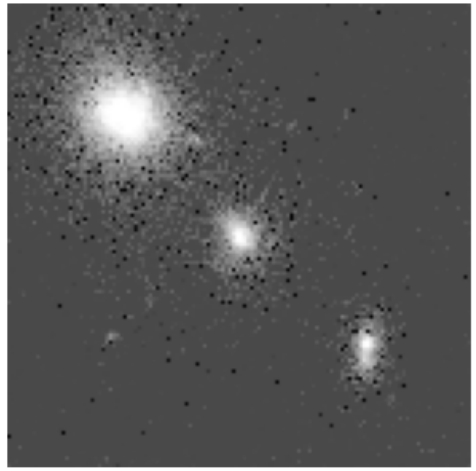
MCG draw



DPS draw



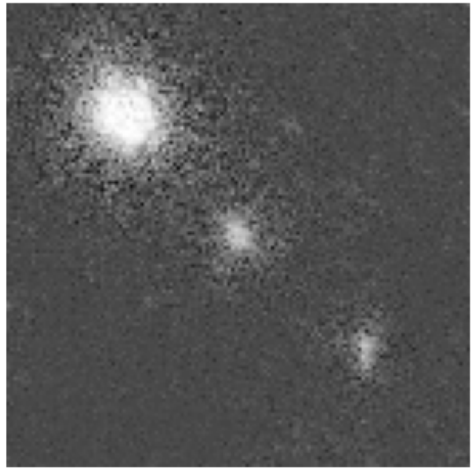
DDRM draw



MCG mean

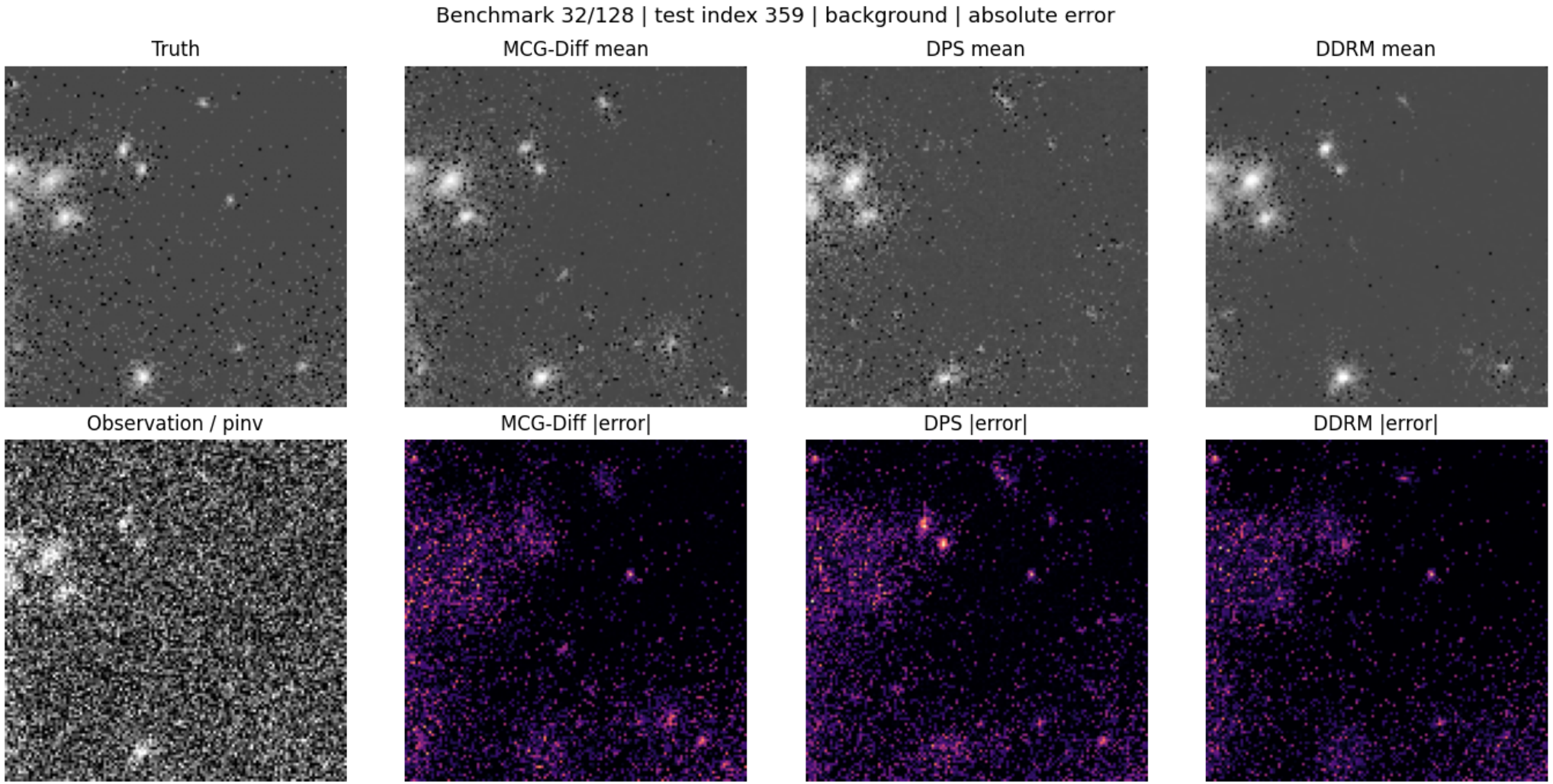


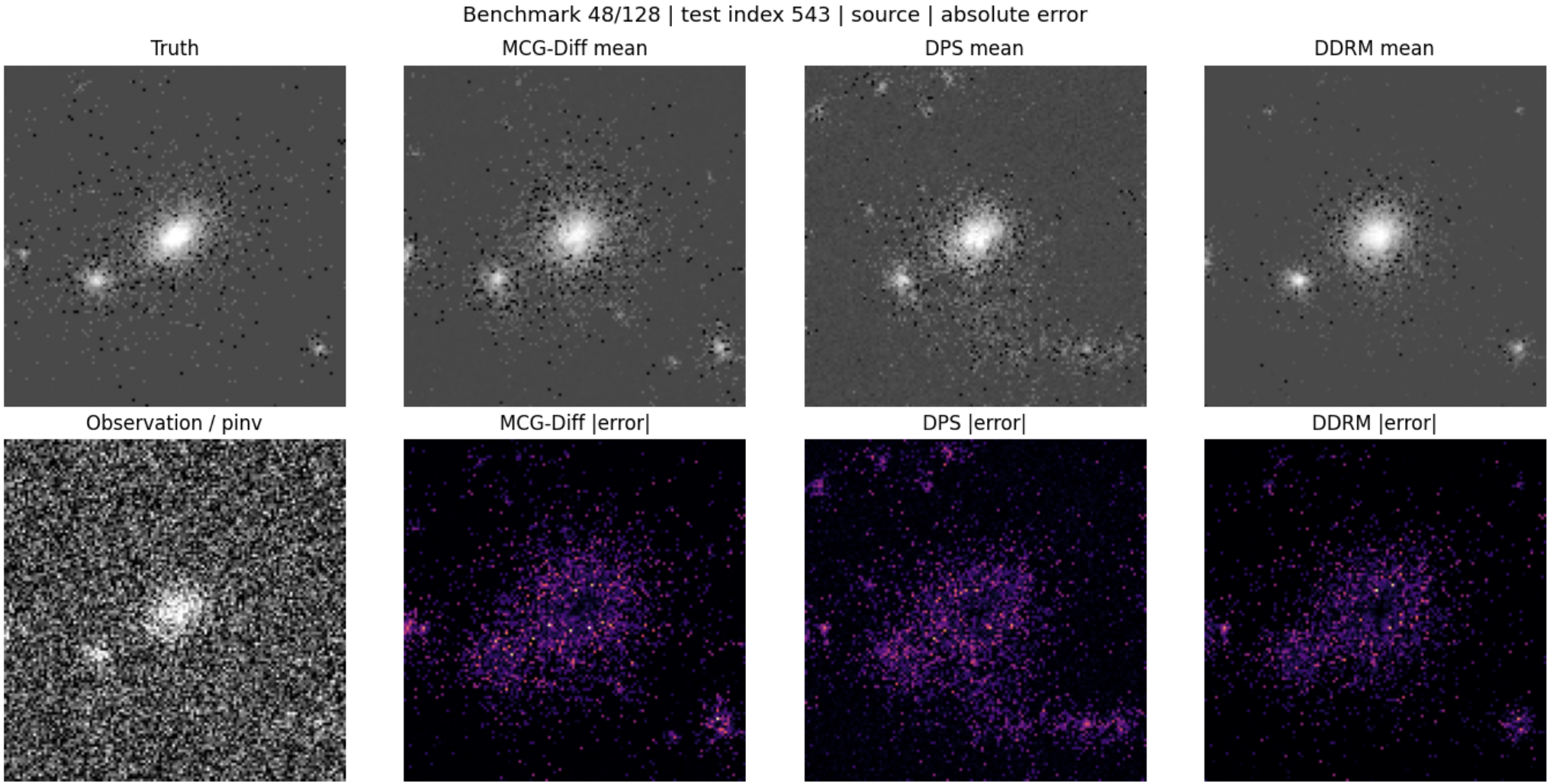
DPS mean

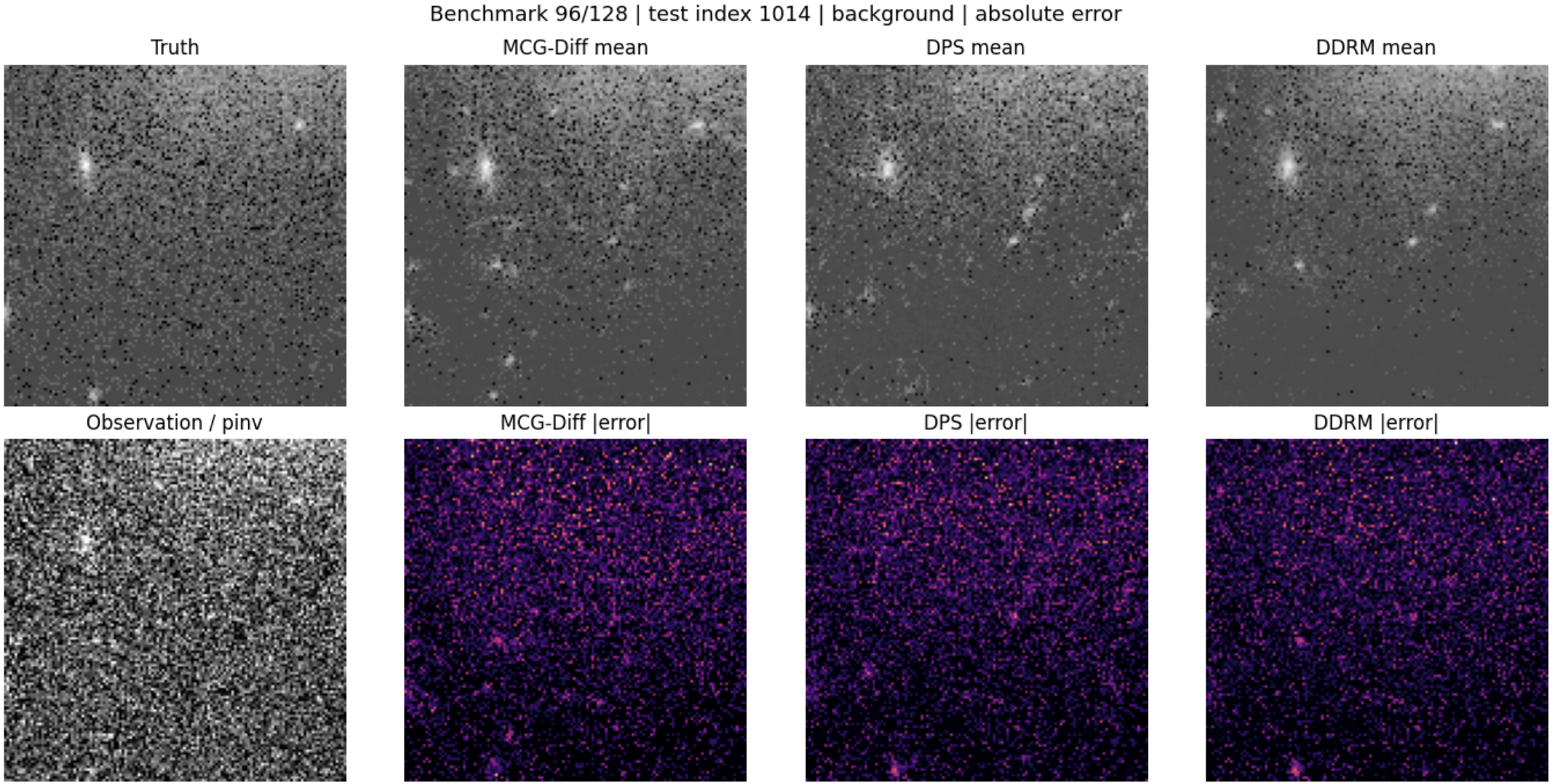


DDRM mean

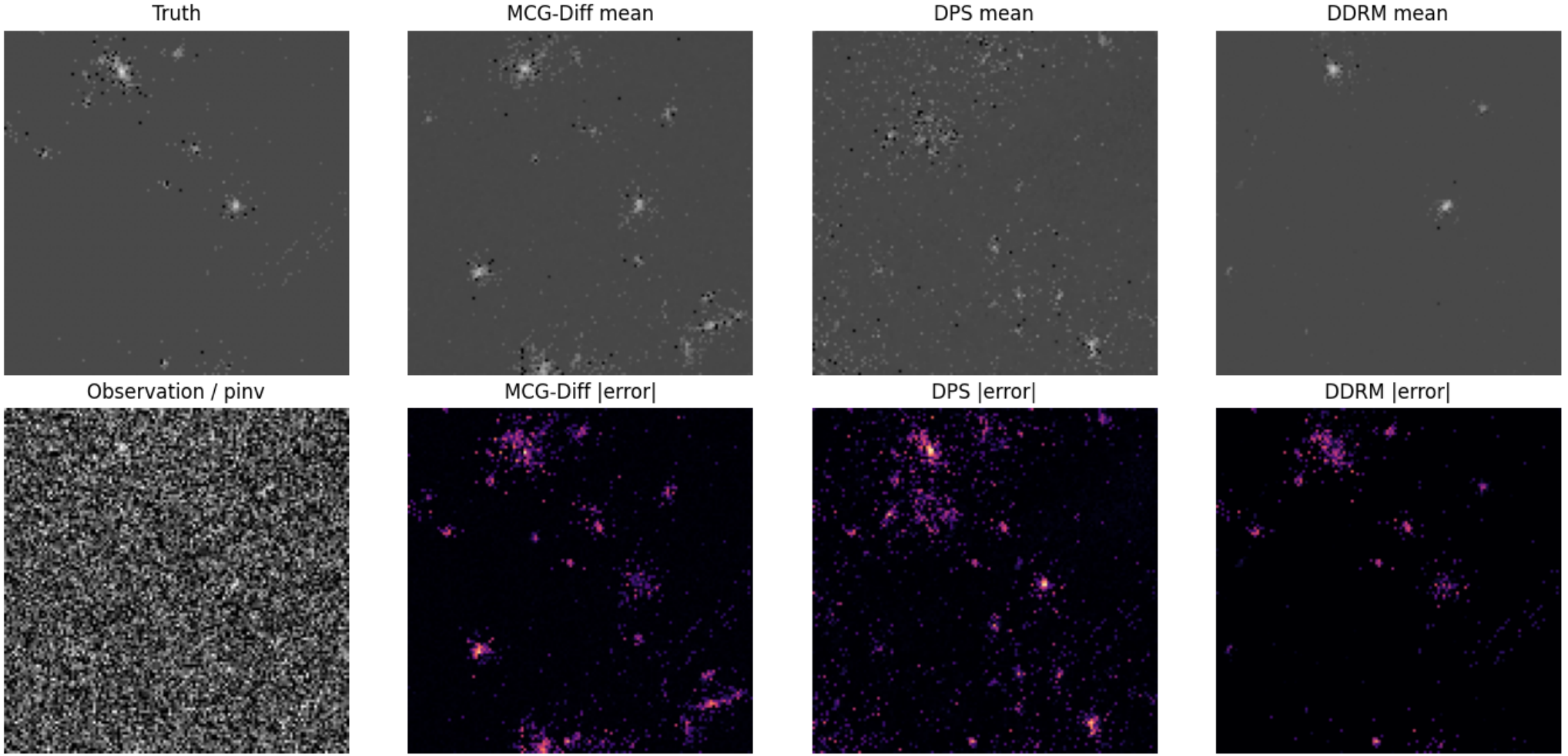


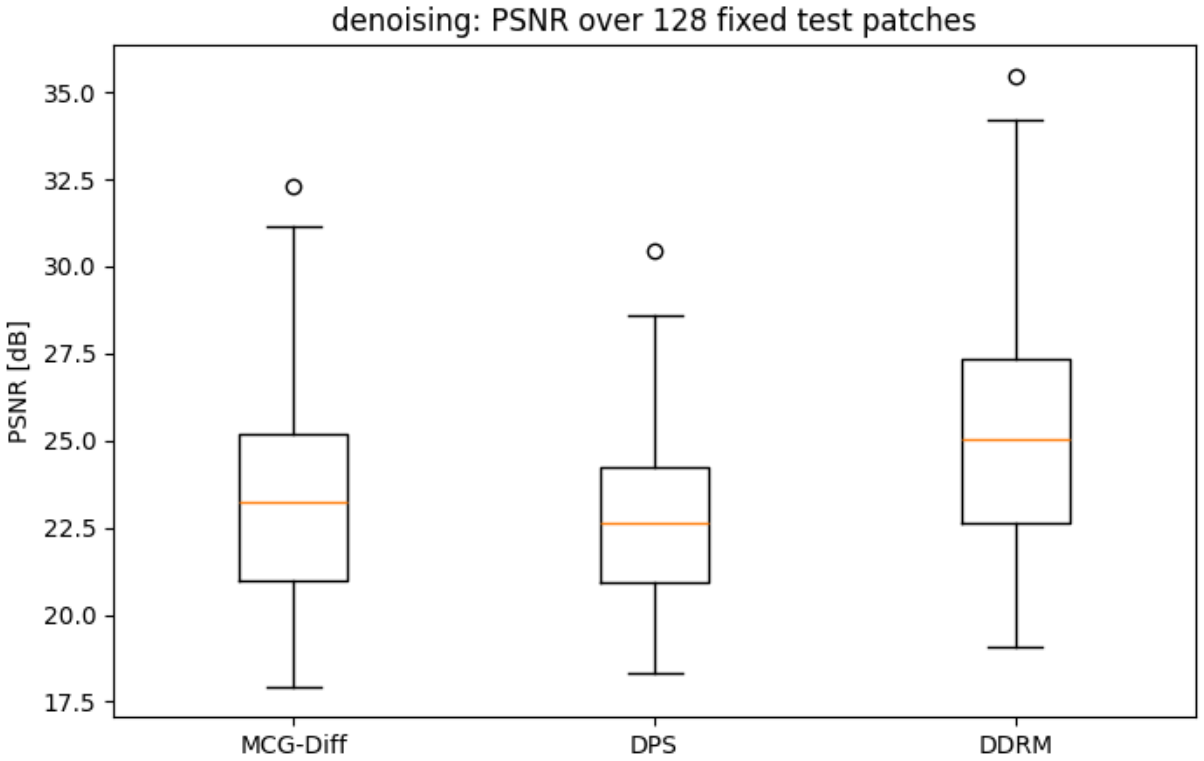

Benchmark 32/128 | test index 359 | background | absolute error
Truth
MCG-Diff mean
DPS mean
DDRM mean
Observation / pinv
MCG-Diff |error|
DPS |error|
DDRM |error|

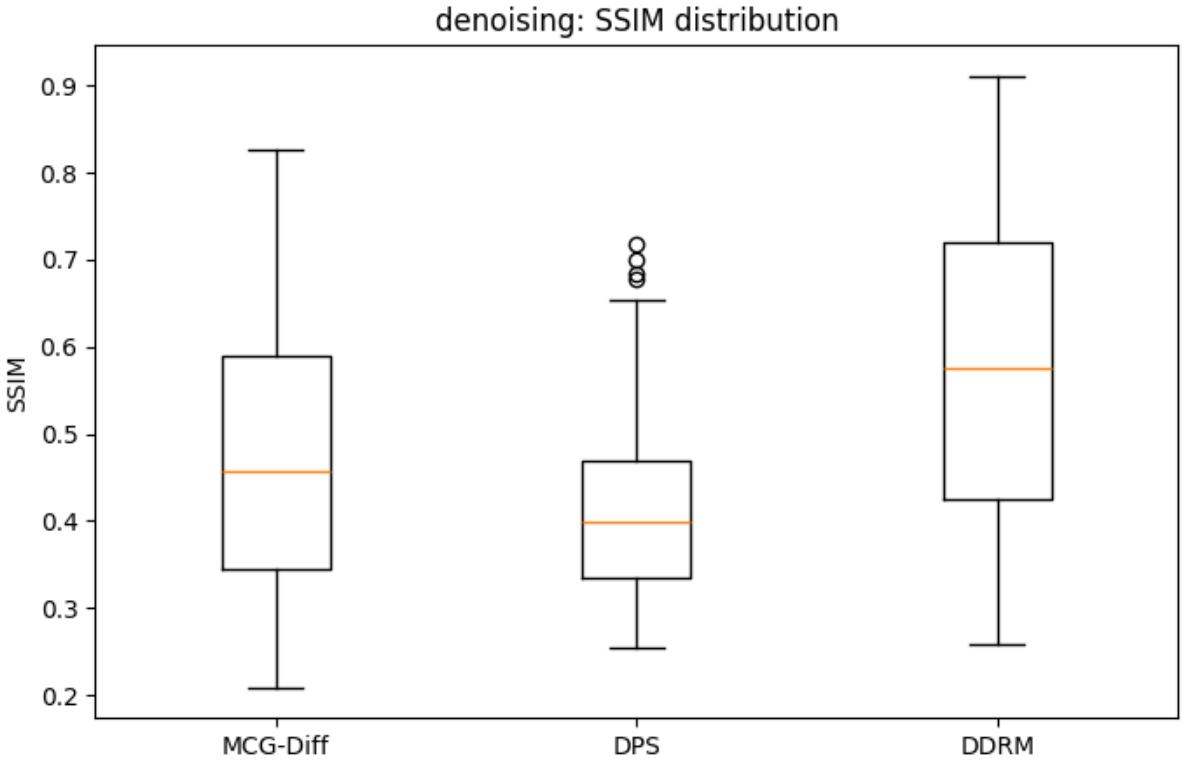

Benchmark 48/128 | test index 543 | source | absolute error
Truth
MCG-Diff mean
DPS mean
DDRM mean
Observation / pinv
MCG-Diff |error|
DPS |error|
DDRM |error|


Benchmark 96/128 | test index 1014 | background | absolute error
Truth
MCG-Diff mean
DPS mean
DDRM mean
Observation / pinv
MCG-Diff |error|
DPS |error|
DDRM |error|

Benchmark 128/128 | test index 1649 | background | absolute error




denoising: PSNR over 128 fixed test patches
PSNR [dB]
35.0
32.5
30.0
27.5
25.0
22.5
20.0
17.5
MCG-Diff
DPS
DDRM


denoising: SSIM distribution
SSIM
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
MCG-Diff
DPS
DDRM

. . . and finally . . .

- ...and beyond!

  Pretty much the same Langevin-based paradigm can be adapted to deal with a variety of **inverse problems** in order to recover a signal from **corrupted measurements**:

$$y = \mathrm{H}x + \epsilon$$

  - De-blurring
  - In-Painting
  - De-Mosaicing
  - Super-Resolution

- Power is nothing without control



PNRR Activity: Topological Data Analysis for Astronomical Imaging

- ...and beyond!

  Pretty much the same Langevin-based paradigm can be adapted to deal with a variety of **inverse problems** in order to recover a signal from **corrupted measurements**:

$$y = \mathsf{H}x + \epsilon$$

  - De-blurring
  - In-Painting
  - De-Mosaicing
  - Super-Resolution

- Power is nothing without control
  - Often we'd to ensure the generated data not only looks realistic but also adheres to desired **connectivity** constraints.
  - With **topological steering** we guide the denoising process to produce outputs with specific structural properties (e.g. connected components, holes, or specific branching structures)
  - Typical tools are **persistent homology** or **topological loss** functions, although they are quite **computationally intensive**, **but...**

Finanziato dall'Unione europea
NextGenerationEU

Ministero dell'Università e della Ricerca

Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA

INAF
ISTITUTO NAZIONALE DI ASTROFISICA

Ric Strikes back!



- Persistent homology is expensive on large images.
- Complexes grow quickly in **time** and **memory**.
- Standard pipelines are impractical for real surveys.
- PixHomology makes PH feasible on real images.

PixHomology

https://github.com/riccardoc95/PixHomology

PNRR
Missione 4 • Componente 2
Investimento 3.1

STILES – IR0000034
CUP C33C22000640006

Finanziato dall'Unione europea
NextGenerationEU

Ministero dell'Università e della Ricerca

Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA

INAF
ISTITUTO NAZIONALE DI ASTROFISICA

Ric Strikes back!



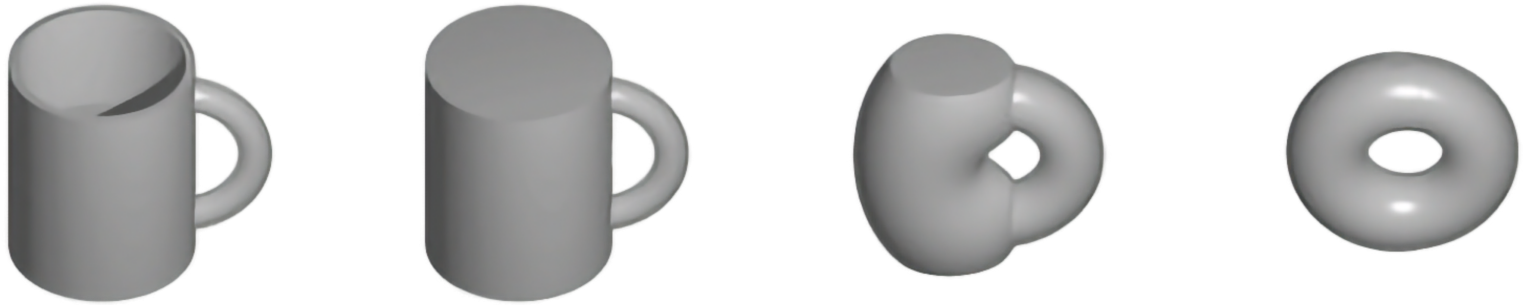
- Persistent homology is expensive on large images.
- C…me and memory.
- S…ctical for real surveys.
- P…ible on real images.

PixHomology

https://github.com/riccardoc95/PixHomology

What is Topology

saddest mathematical joke of all times



Topology is a way of investigating and representing the connectivity structure of an object.

- 0 dimension: **connected component**
- 1 dimension: **loops**
- (…)

Why do we need Topology

in data analysis in general



Topology provides a good characterization of the shape of the data that grants:

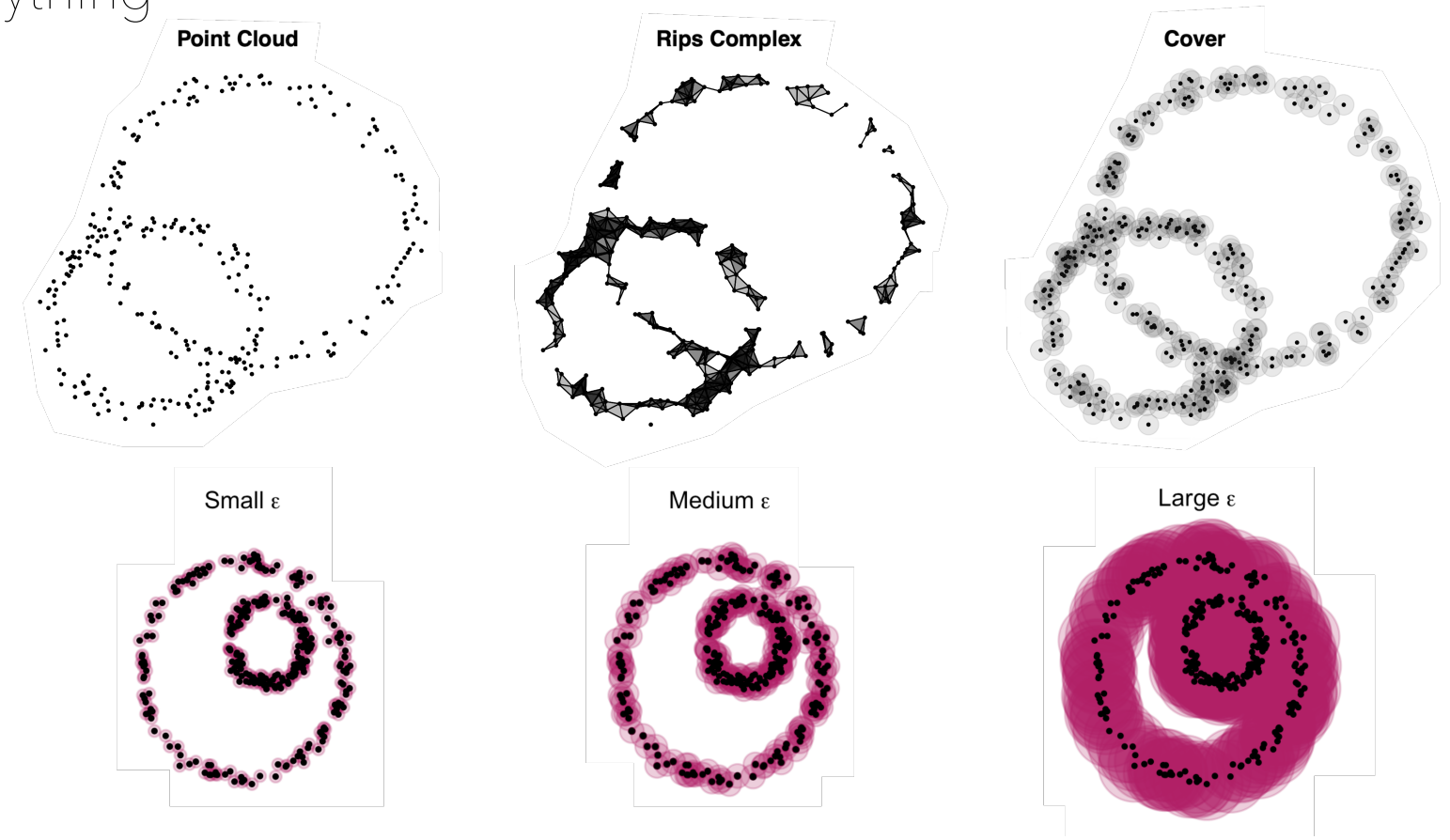
- Coordinate Invariance
- Deformation Invariance
- Easy Interpretation.

Source: Carlsson, G. (2009). Topology and data. Bulletin of the AMS, 46(2), 255-308

Topological Data Analysis

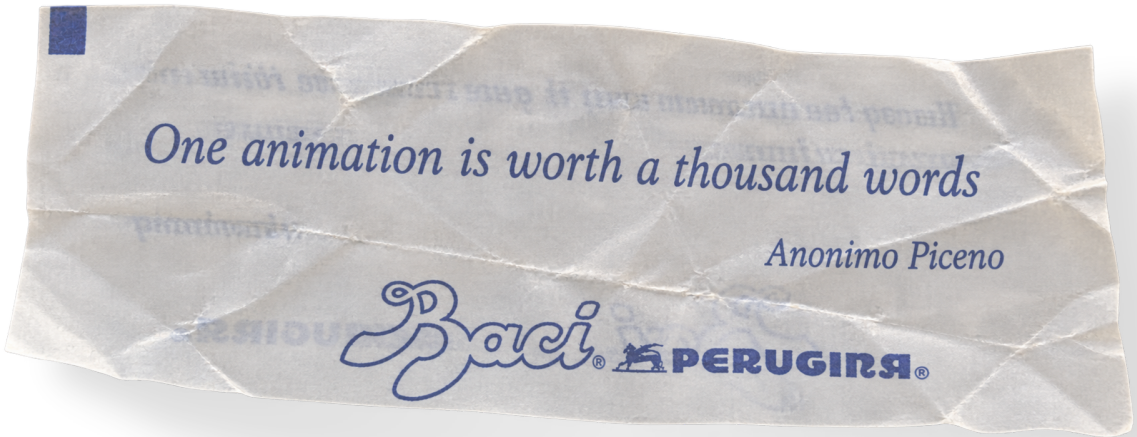
a filtration for everything

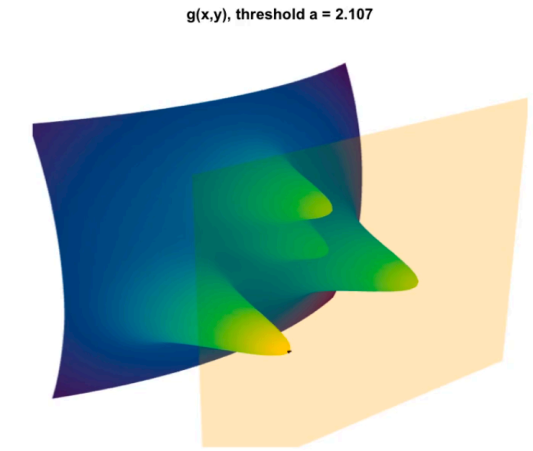
↬ POINT CLOUD ↫

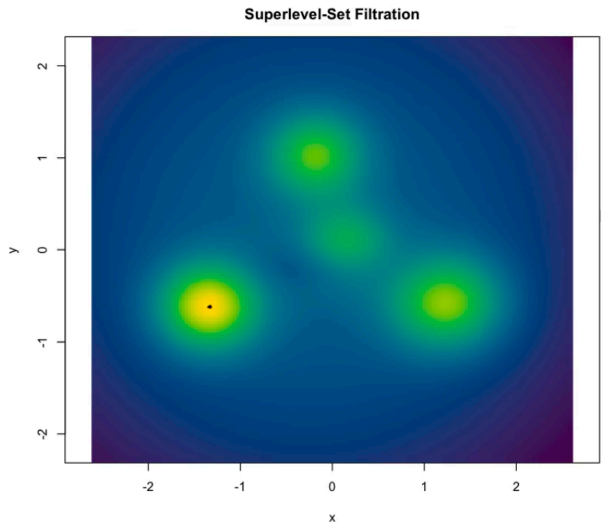


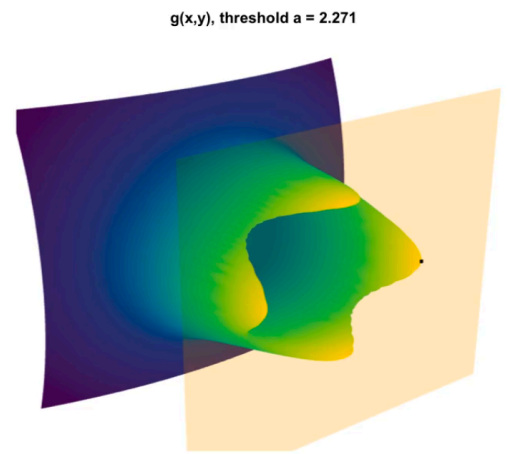
Topological Data Analysis

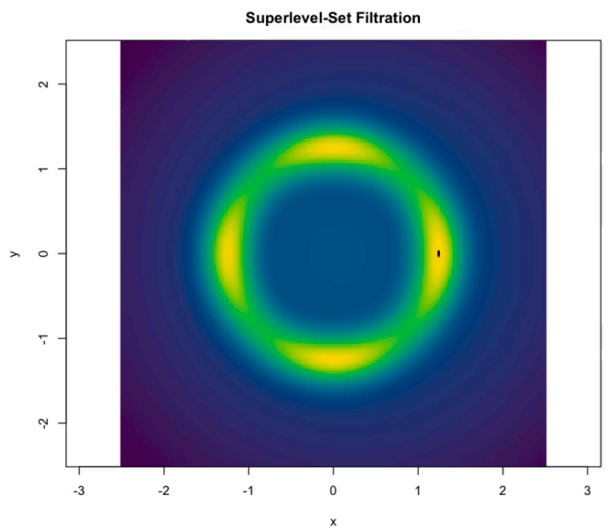
a filtration for everything

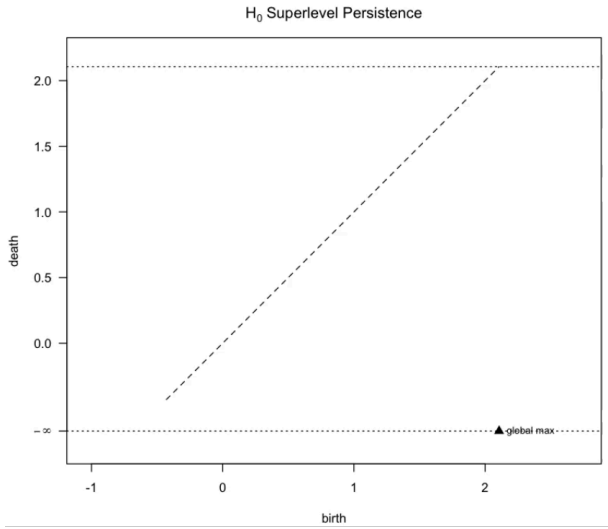



g(x,y), threshold a = 2.107

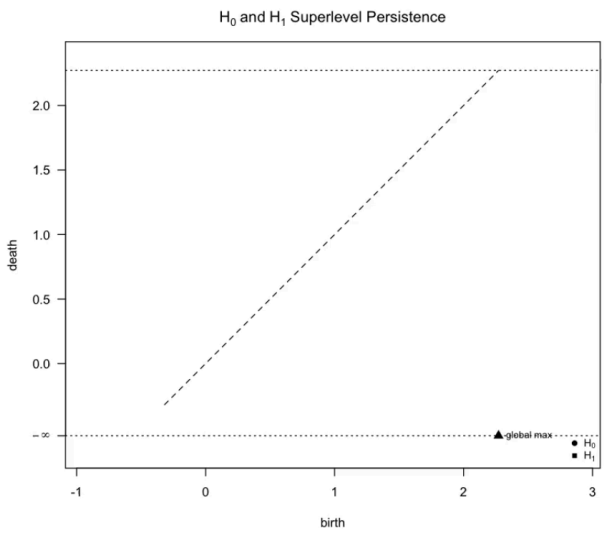

Superlevel-Set Filtration
x
y

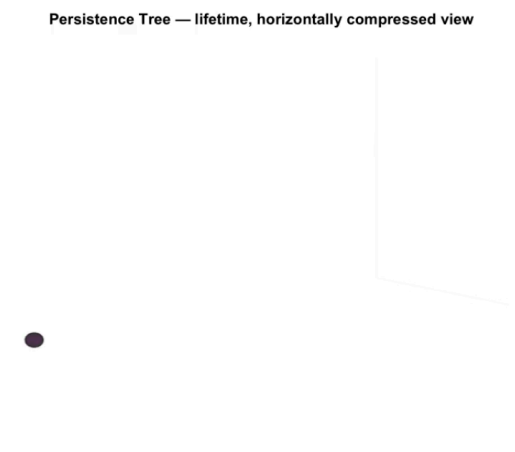

g(x,y), threshold a = 2.271

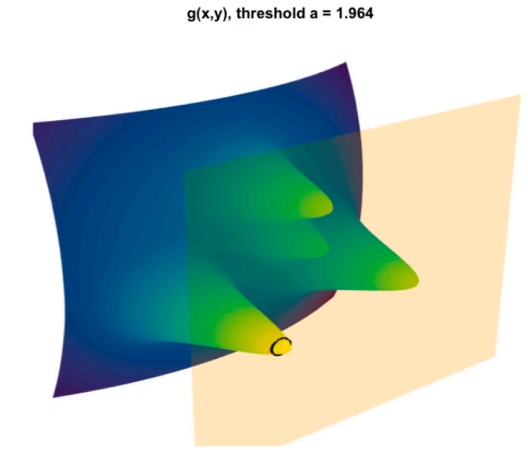

Superlevel-Set Filtration
x
y

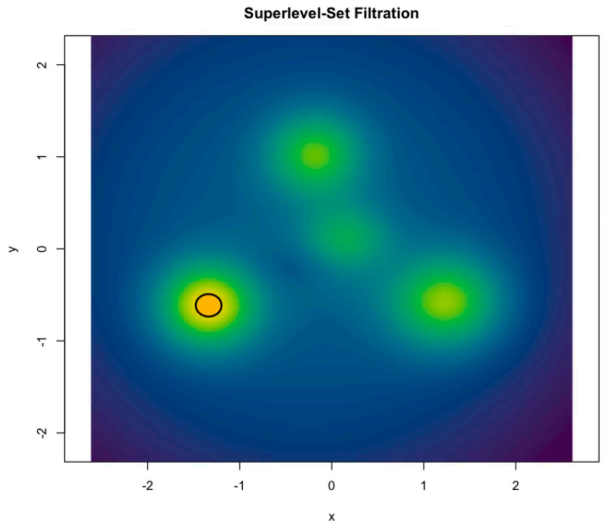

H0 Superlevel Persistence
death
birth
global max

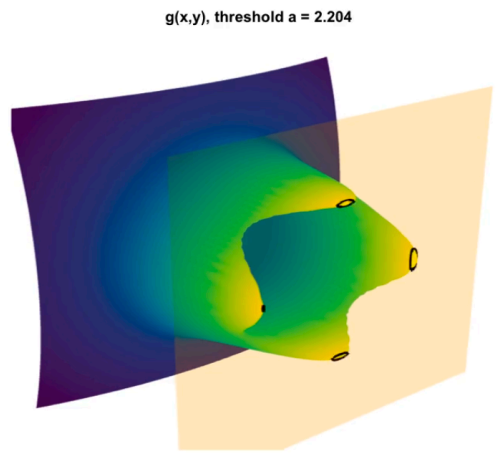

Persistence Tree — lifetime, right-to-left view

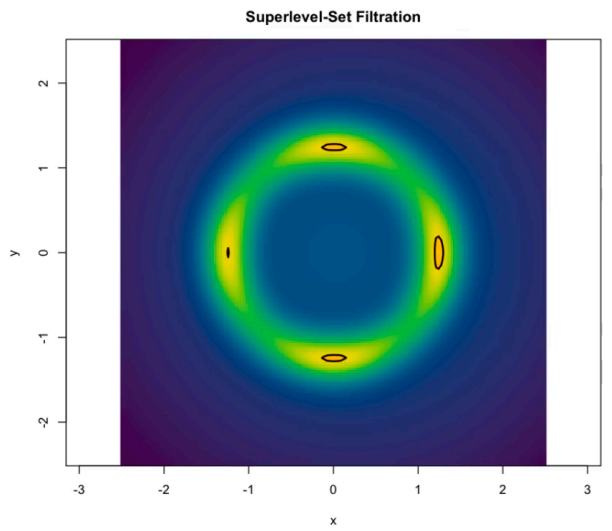

H0 and H1 Superlevel Persistence
death
birth
global max
H0
H1

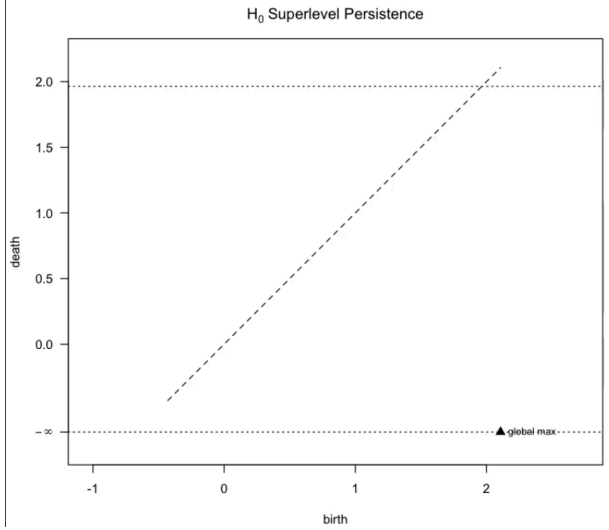

Persistence Tree — lifetime, horizontally compressed view


g(x,y), threshold a = 1.964

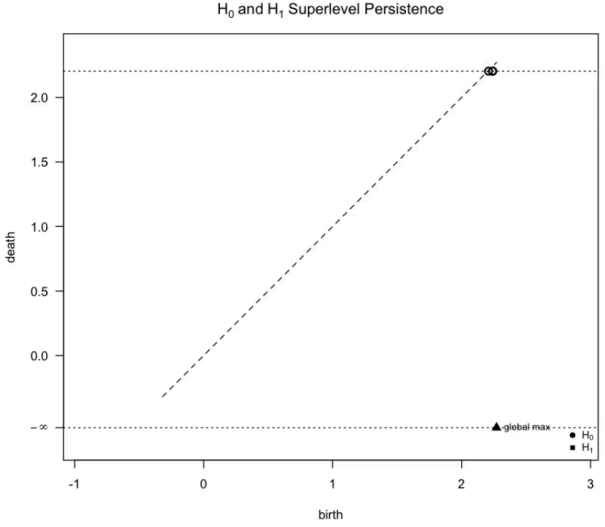

Superlevel-Set Filtration
y
x


g(x,y), threshold a = 2.204

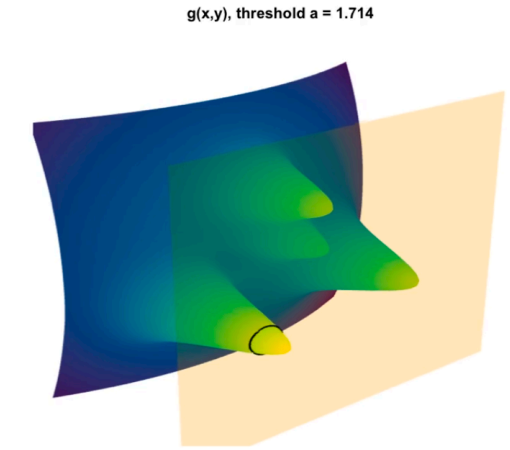

Superlevel-Set Filtration
y
x

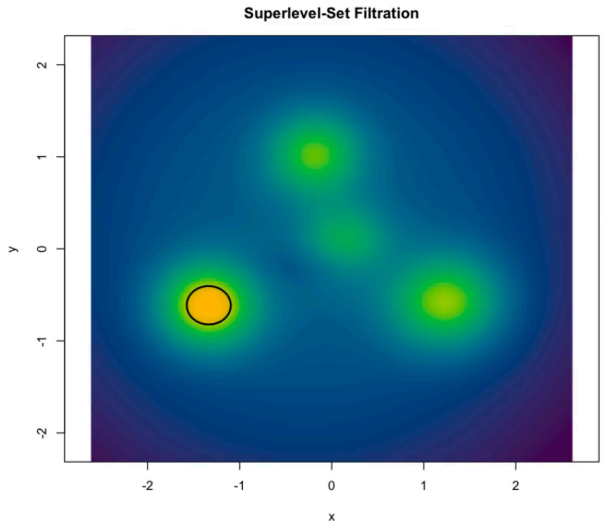

H0 Superlevel Persistence
death
birth
global max

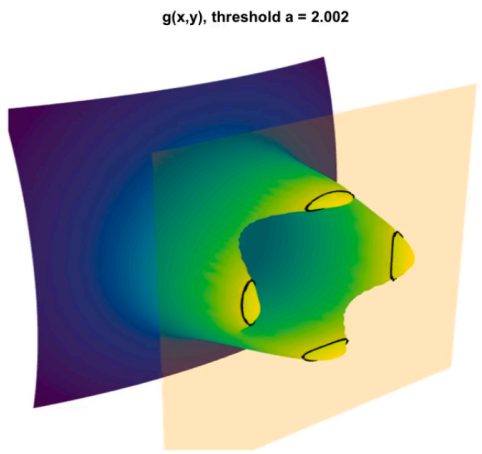

Persistence Tree — lifetime, right-to-left view

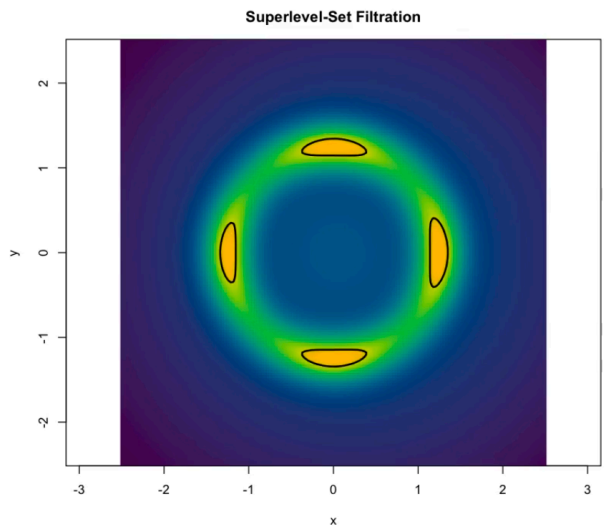

H0 and H1 Superlevel Persistence
death
birth
global max
H0
H1

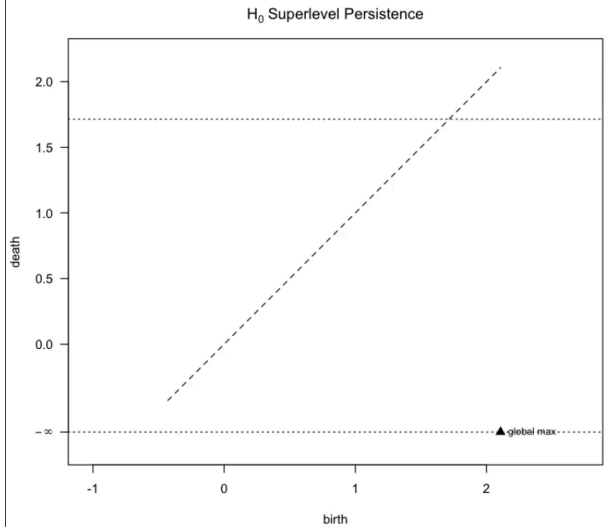

Persistence Tree — lifetime, horizontally compressed view


g(x,y), threshold a = 1.714

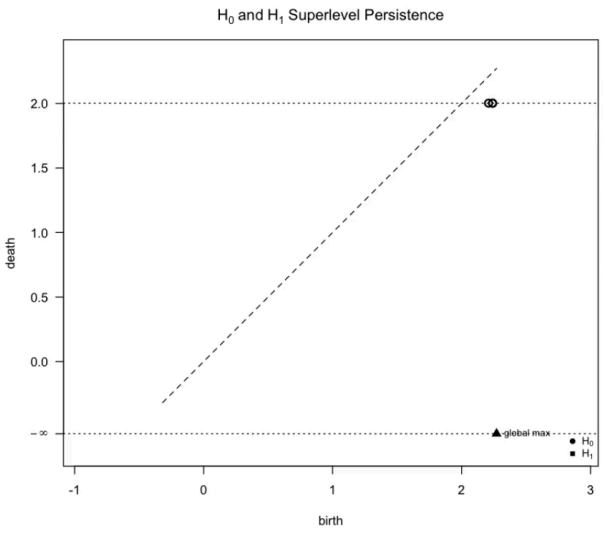

Superlevel-Set Filtration
x
y

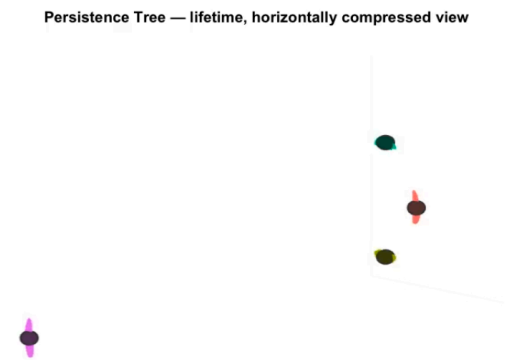

g(x,y), threshold a = 2.002

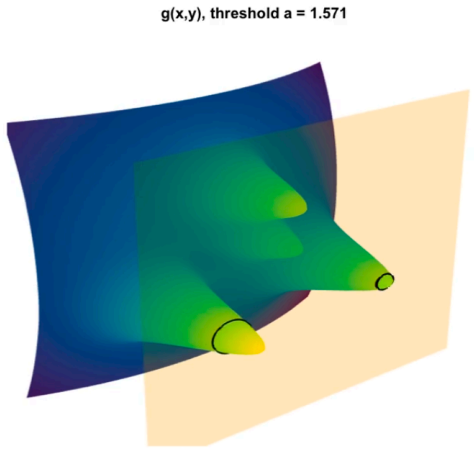

Superlevel-Set Filtration
x
y

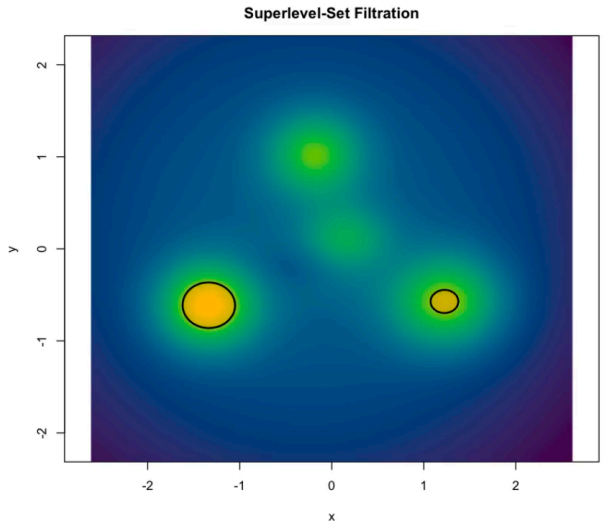

H0 Superlevel Persistence
death
birth
global max

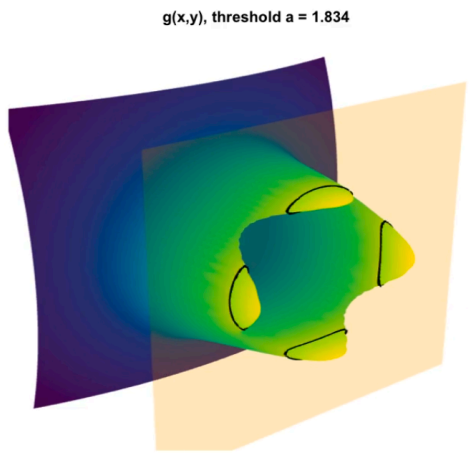

Persistence Tree — lifetime, right-to-left view

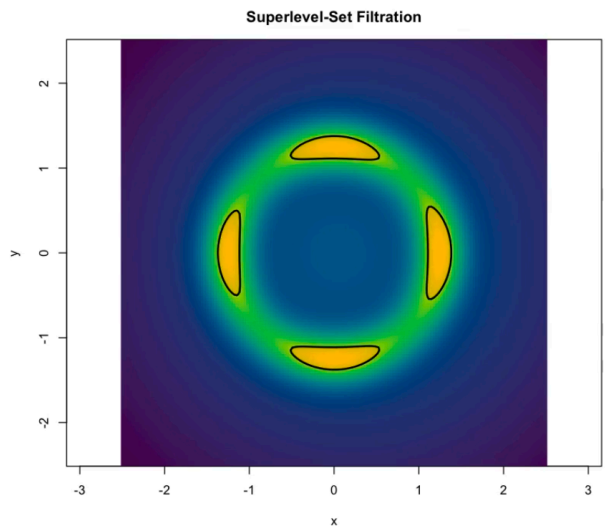

H0 and H1 Superlevel Persistence
death
birth
global max
H0
H1

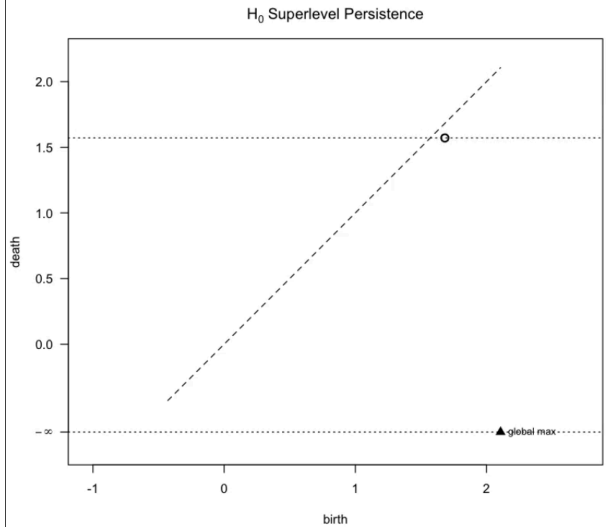

Persistence Tree — lifetime, horizontally compressed view

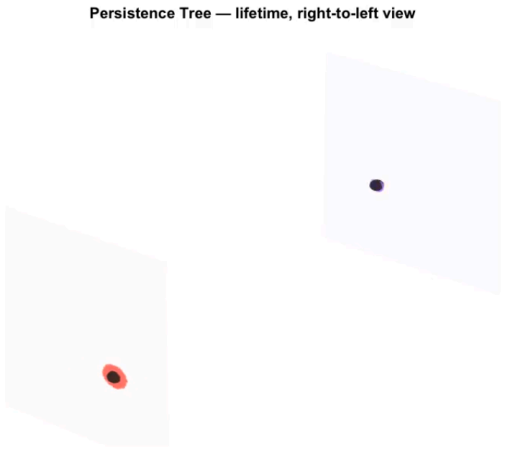

g(x,y), threshold a = 1.571

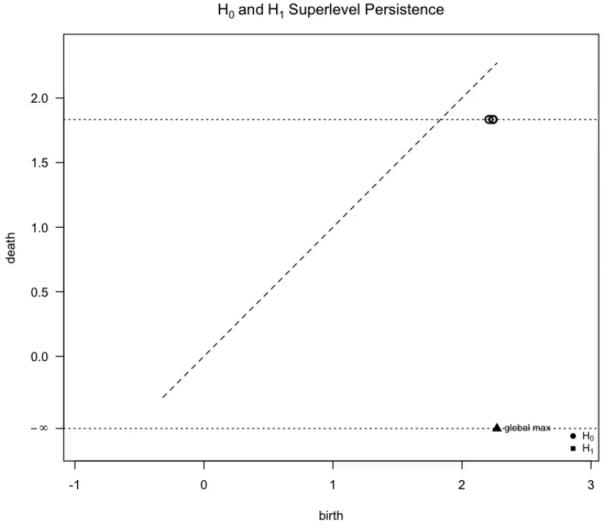

Superlevel-Set Filtration
x
y

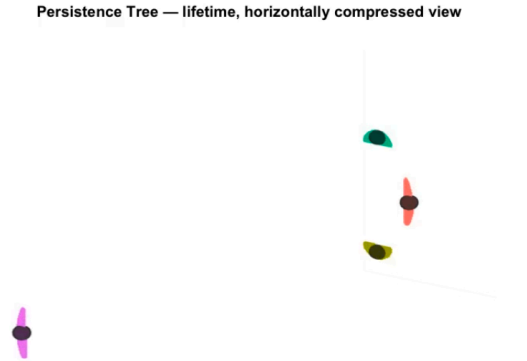

g(x,y), threshold a = 1.834

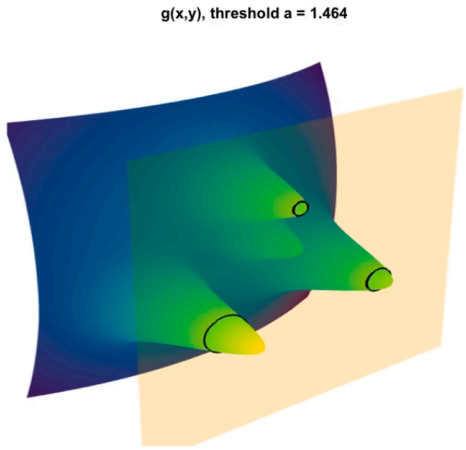

Superlevel-Set Filtration
x
y

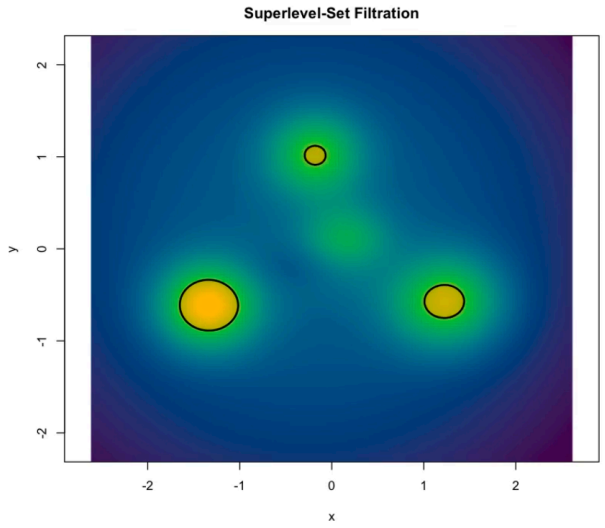

H0 Superlevel Persistence
death
birth
global max

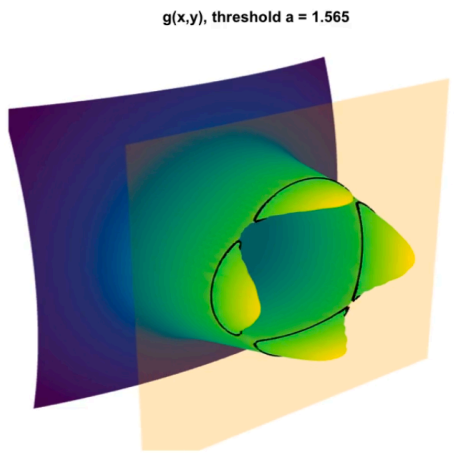

Persistence Tree — lifetime, right-to-left view

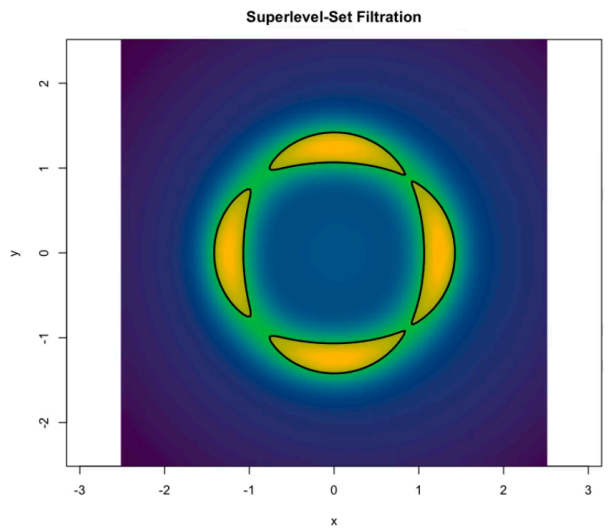

H0 and H1 Superlevel Persistence
death
birth
global max
H0
H1

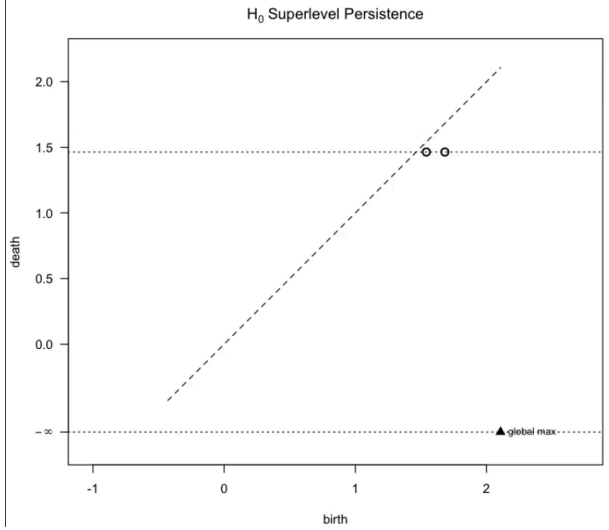

Persistence Tree — lifetime, horizontally compressed view

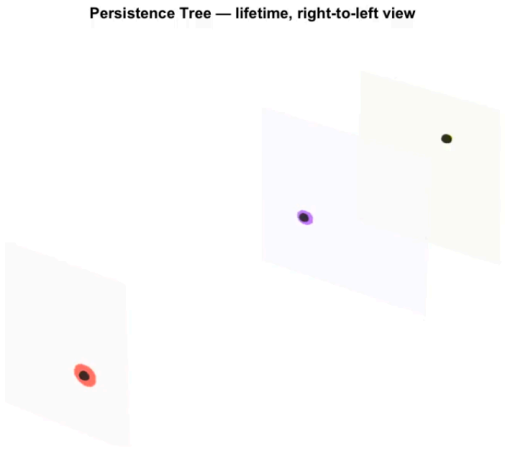

g(x,y), threshold a = 1.464

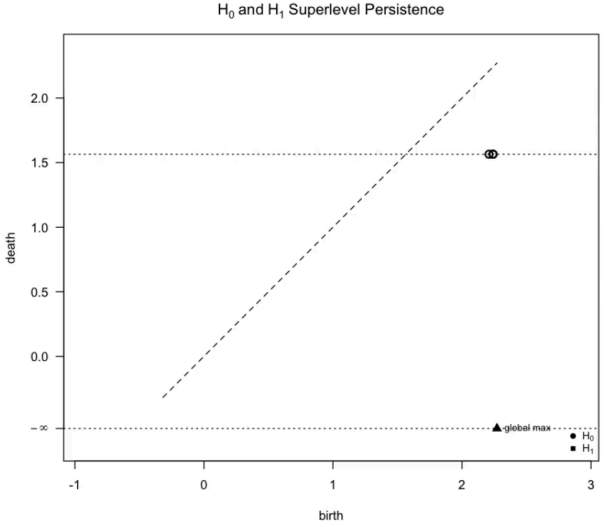

Superlevel-Set Filtration
x
y

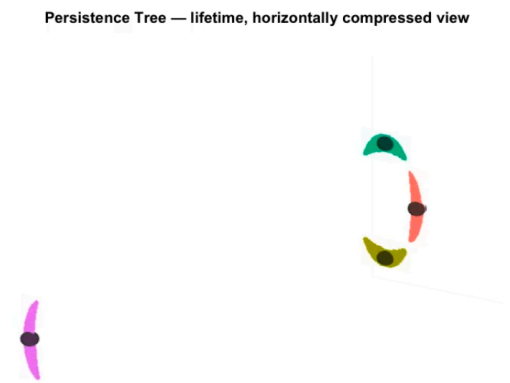

g(x,y), threshold a = 1.565

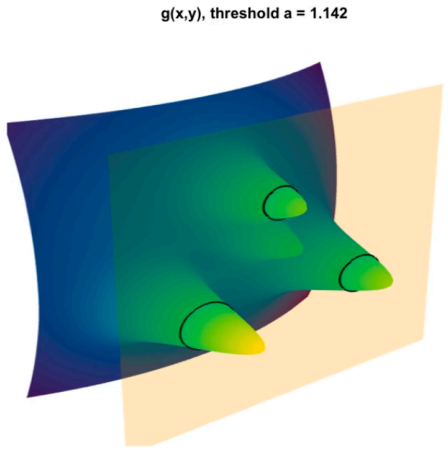

Superlevel-Set Filtration
x
y

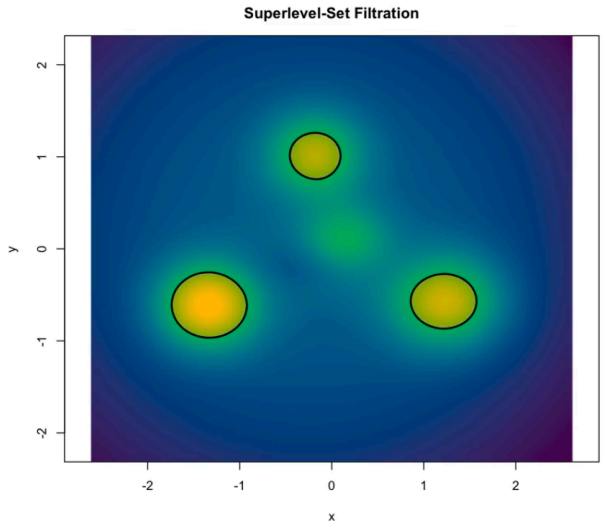

H0 Superlevel Persistence
death
birth
global max

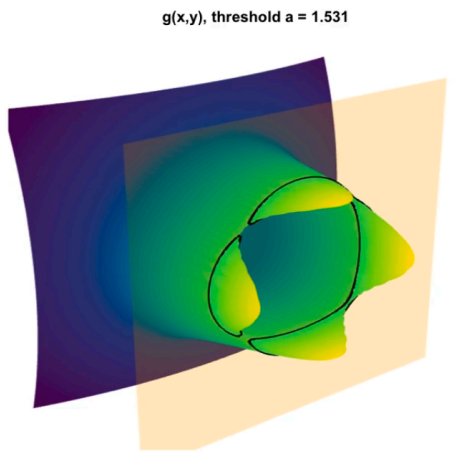

Persistence Tree — lifetime, right-to-left view

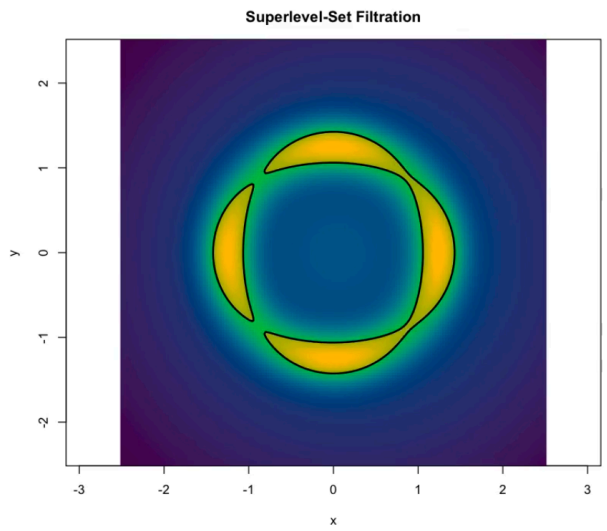

H0 and H1 Superlevel Persistence
death
birth
global max
H0
H1

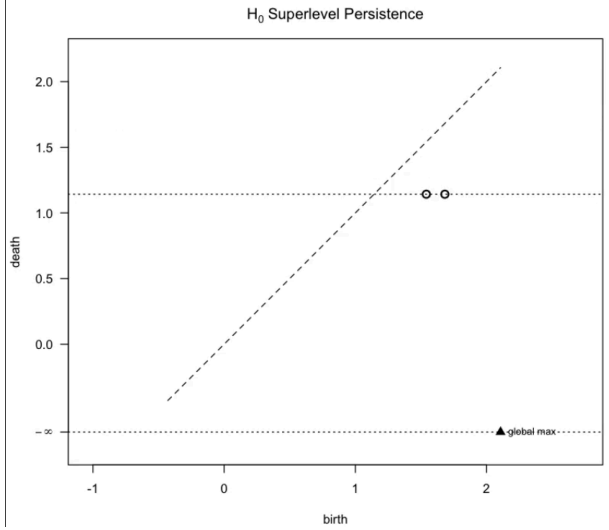

Persistence Tree — lifetime, horizontally compressed view

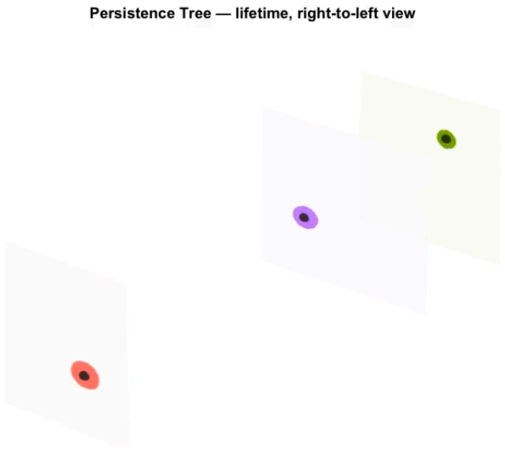

g(x,y), threshold a = 1.142

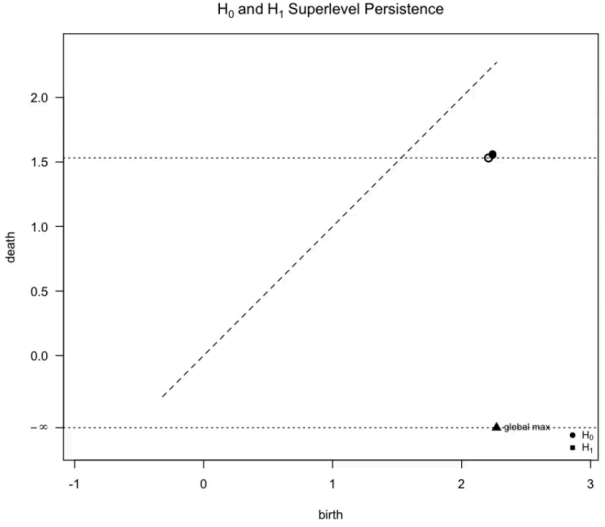

Superlevel-Set Filtration
x
y

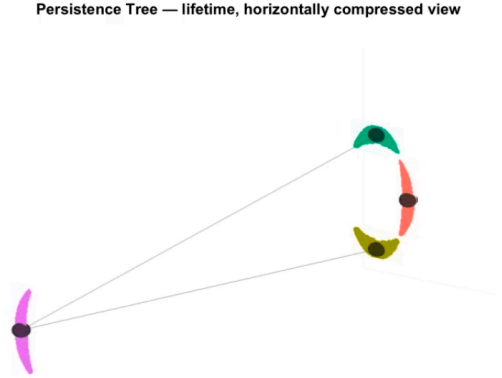

g(x,y), threshold a = 1.531

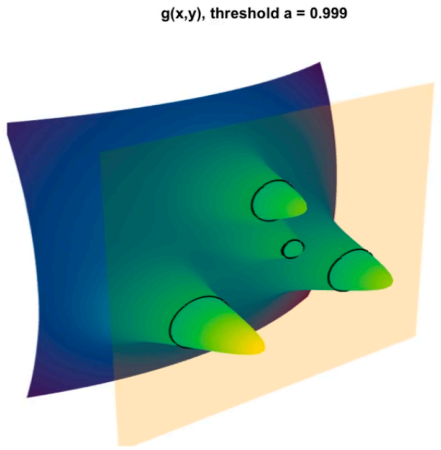

Superlevel-Set Filtration
x
y

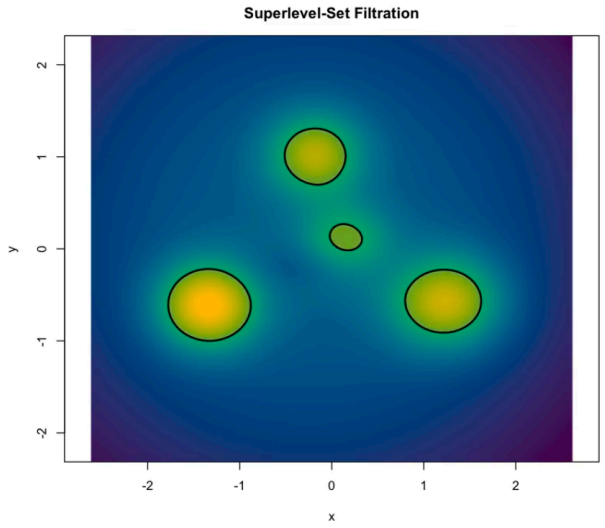

H0 Superlevel Persistence
death
birth
global max

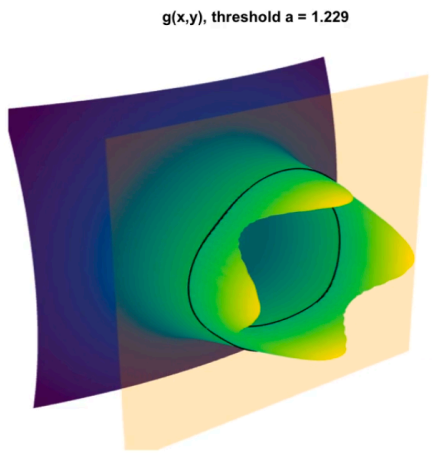

Persistence Tree — lifetime, right-to-left view

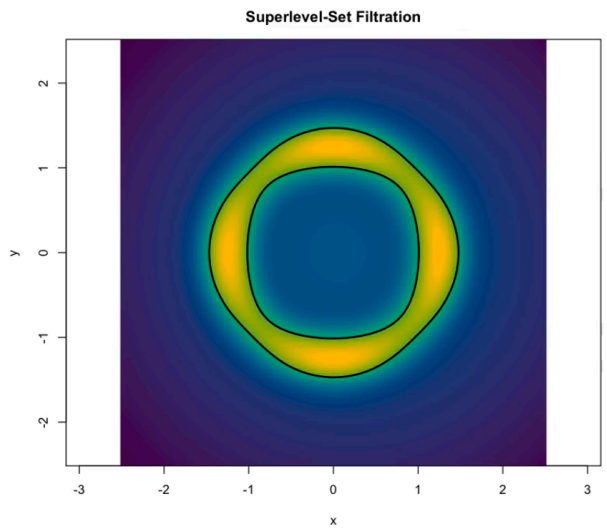

H0 and H1 Superlevel Persistence
death
birth
global max
H0
H1

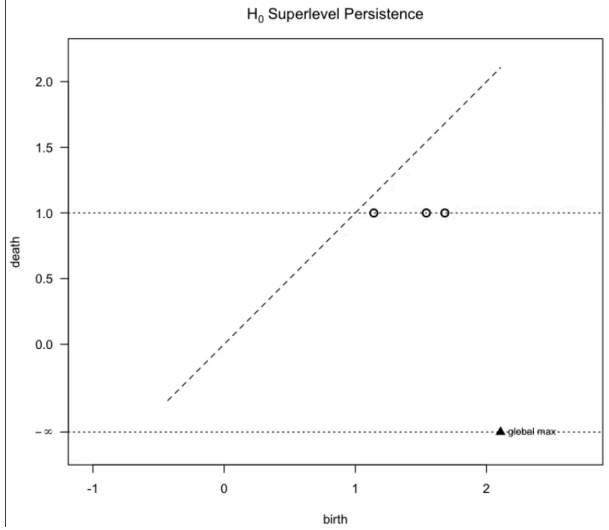

Persistence Tree — lifetime, horizontally compressed view

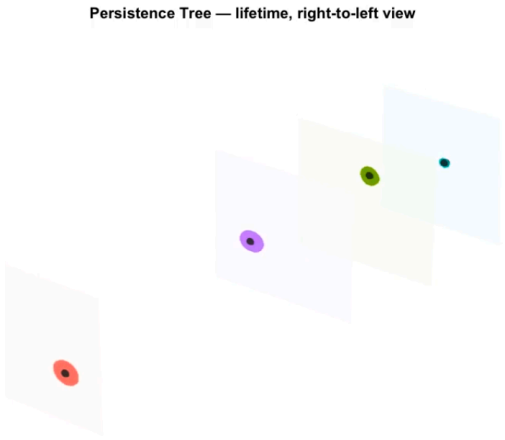

g(x,y), threshold a = 0.999

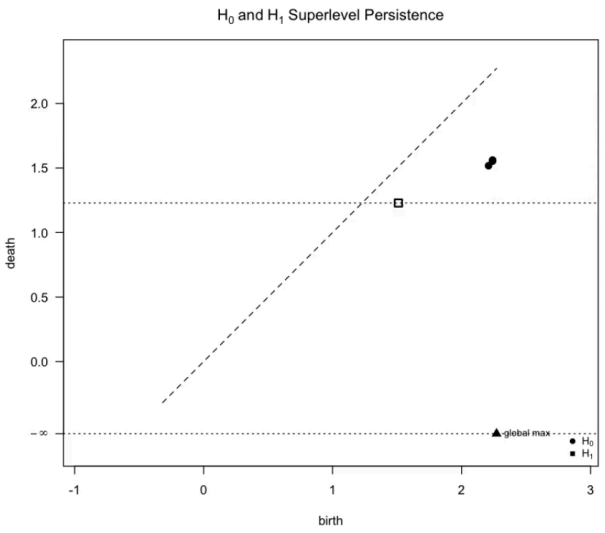

Superlevel-Set Filtration
x
y

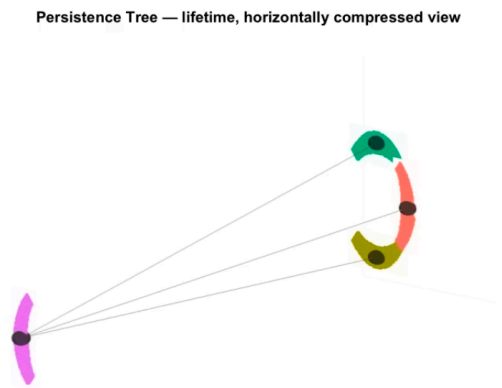

g(x,y), threshold a = 1.229

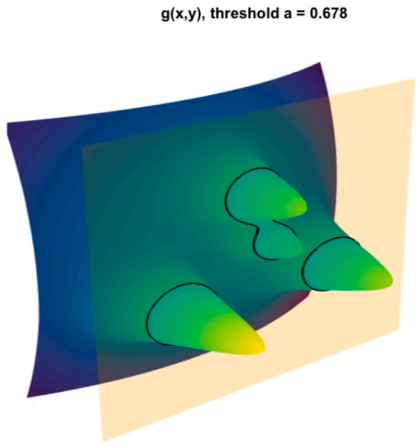

Superlevel-Set Filtration
x
y

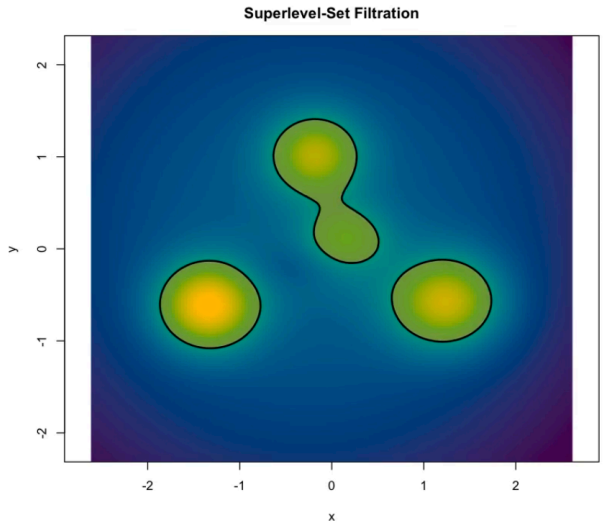

H0 Superlevel Persistence
death
birth
global max

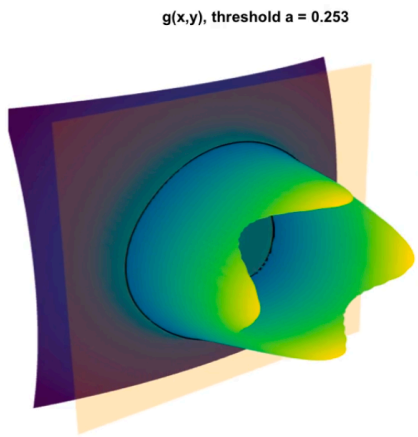

Persistence Tree — lifetime, right-to-left view

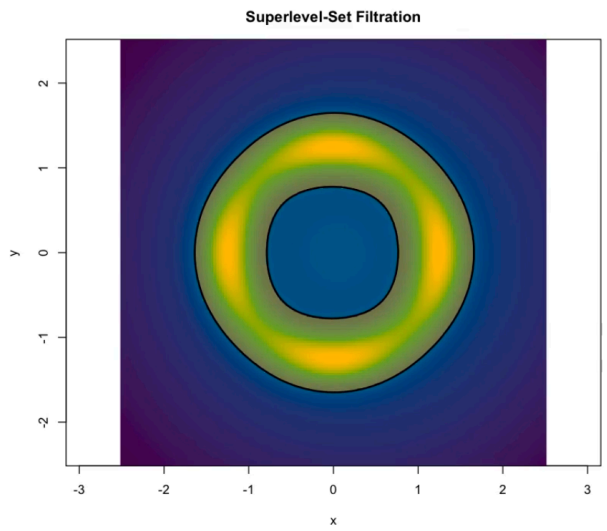

H0 and H1 Superlevel Persistence
death
birth
global max
H0
H1

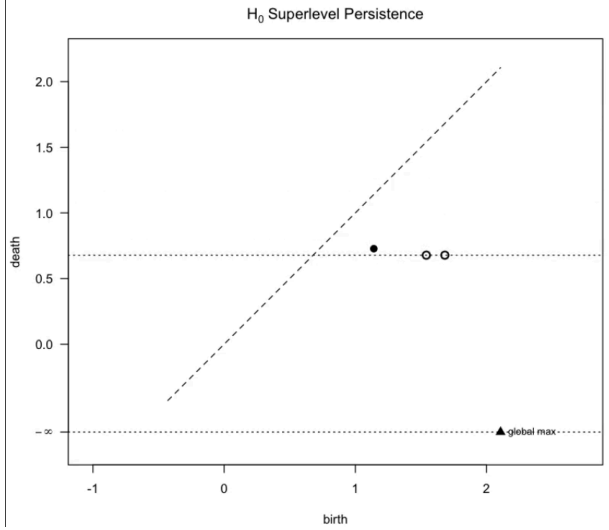

Persistence Tree — lifetime, horizontally compressed view

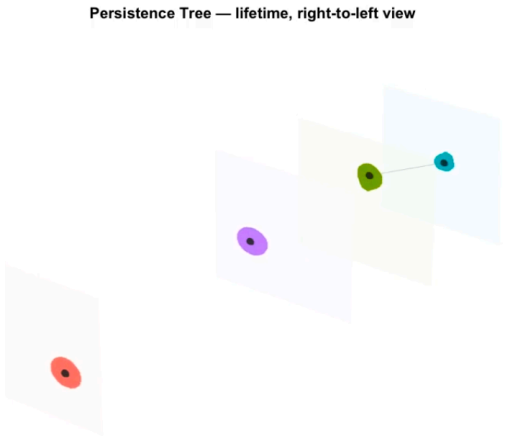

g(x,y), threshold a = 0.678

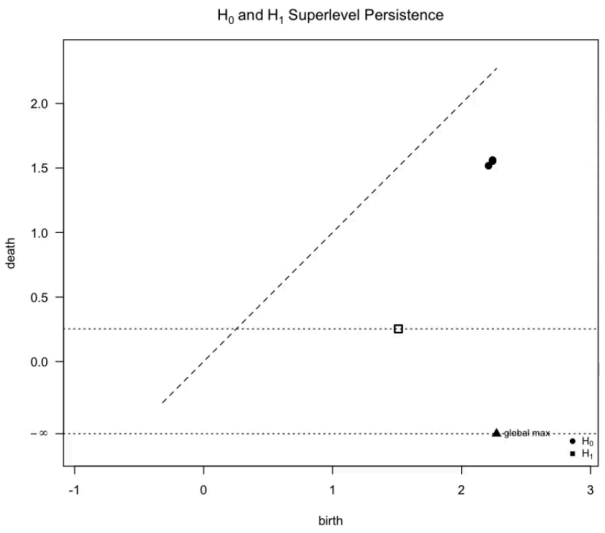

Superlevel-Set Filtration
x
y

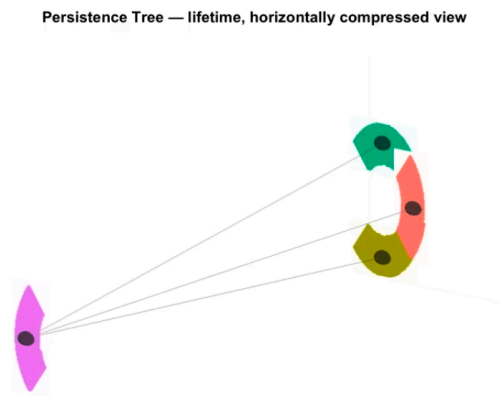

g(x,y), threshold a = 0.253

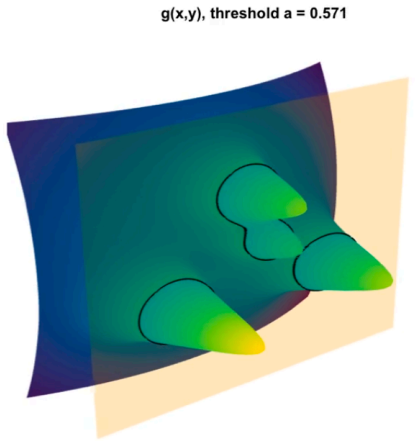

Superlevel-Set Filtration
x
y

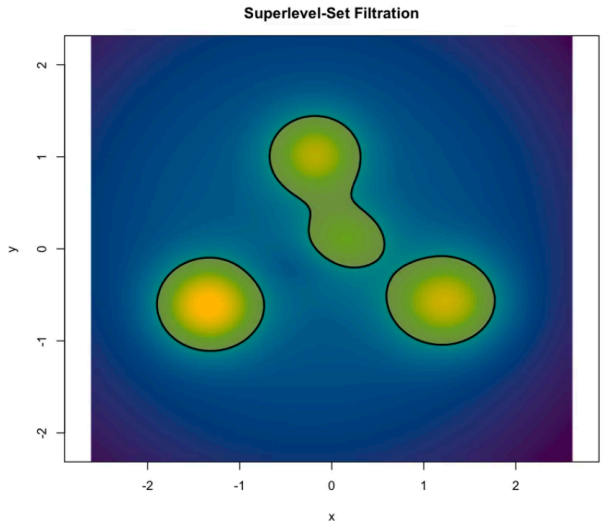

H0 Superlevel Persistence
death
birth
global max

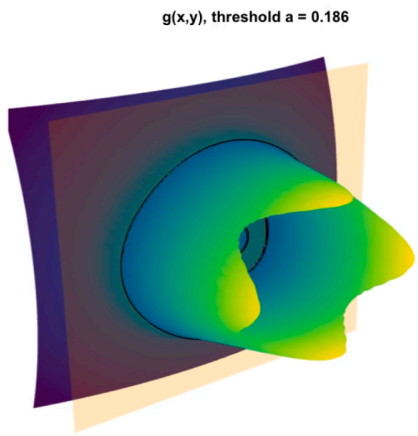

Persistence Tree — lifetime, right-to-left view

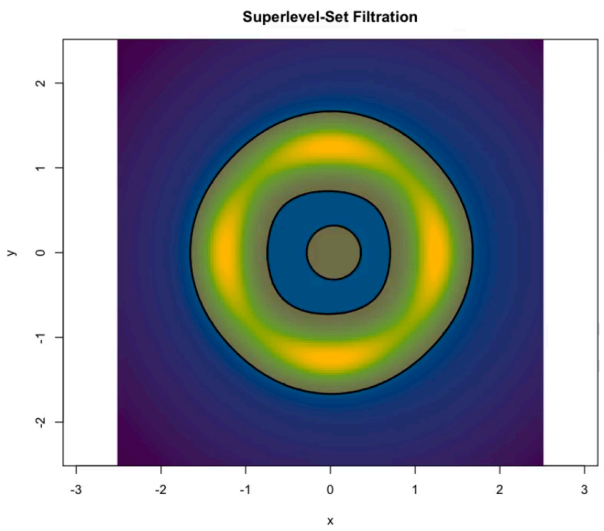

H0 and H1 Superlevel Persistence
death
birth
global max
H0
H1

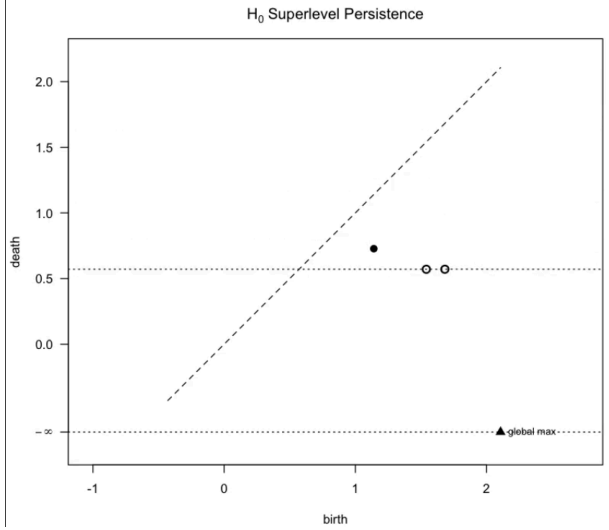

Persistence Tree — lifetime, horizontally compressed view

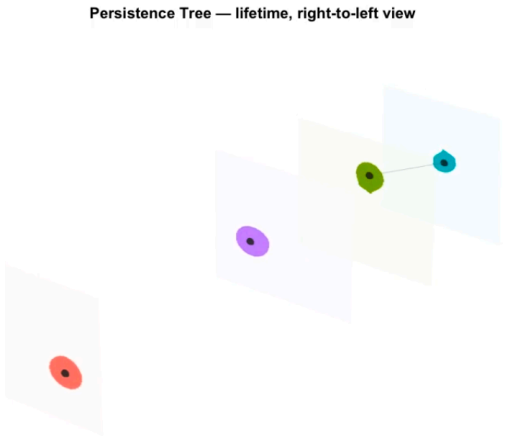

g(x,y), threshold a = 0.571

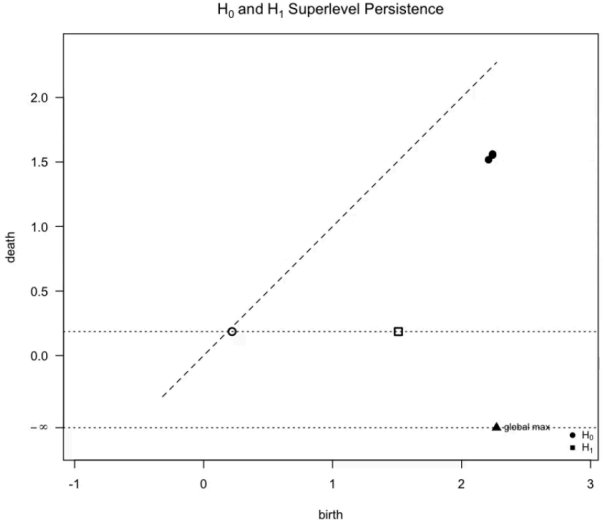

Superlevel-Set Filtration
x
y

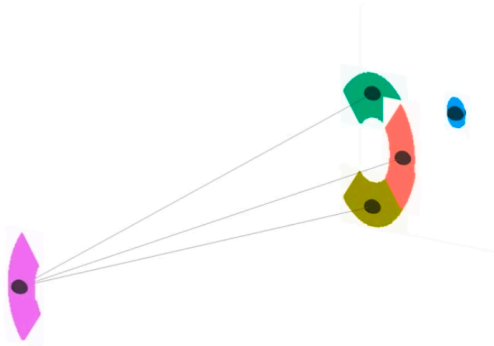

g(x,y), threshold a = 0.186

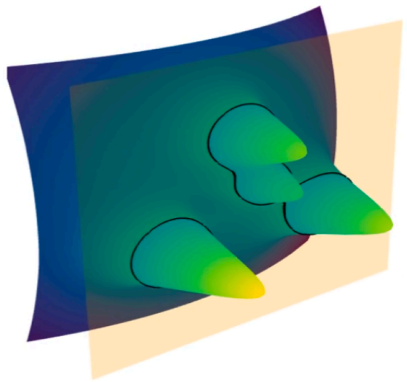

Superlevel-Set Filtration
x
y

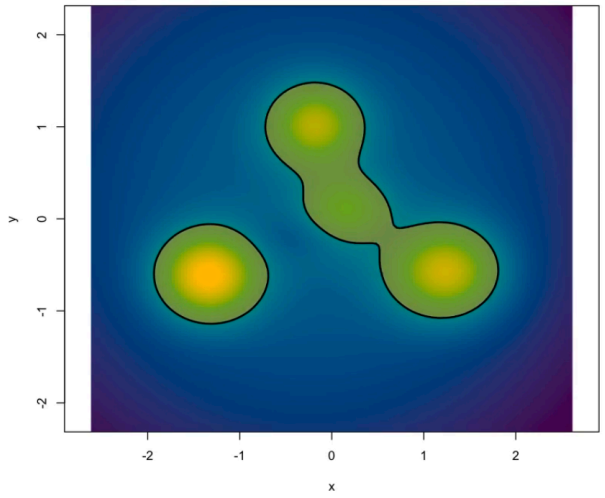

H0 Superlevel Persistence
death
birth
global max

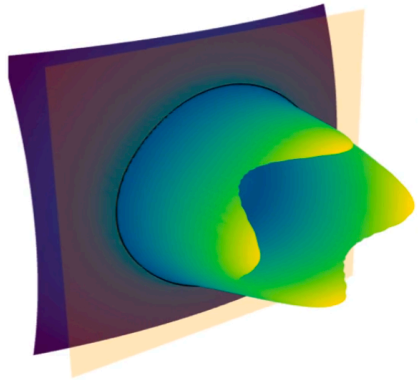

Persistence Tree — lifetime, right-to-left view

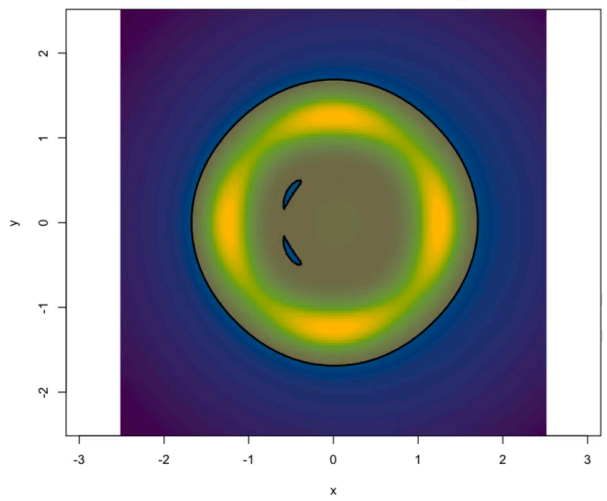

H0 and H1 Superlevel Persistence
death
birth
global max
H0
H1

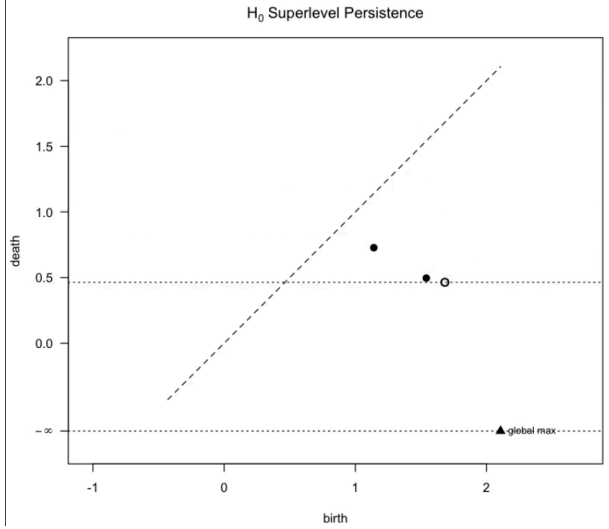

Persistence Tree — lifetime, horizontally compressed view

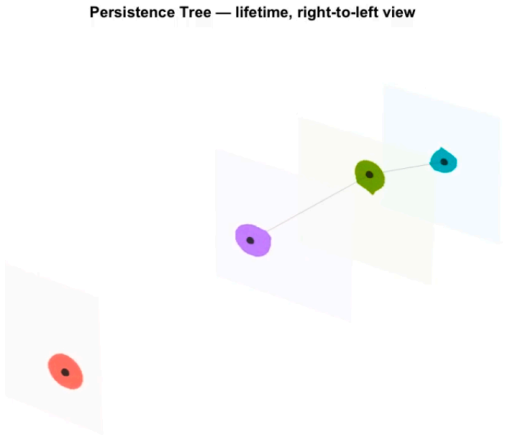

g(x,y), threshold a = 0.464

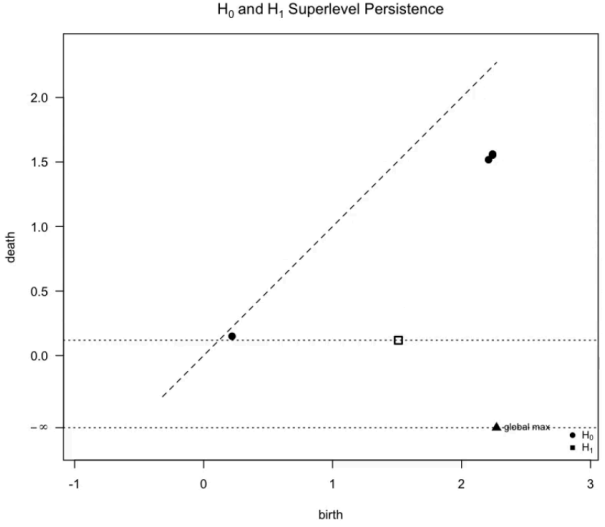

Superlevel-Set Filtration
x
y

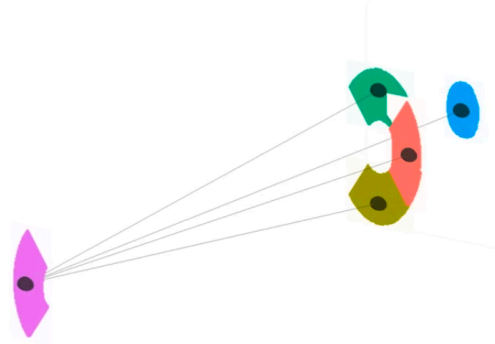

g(x,y), threshold a = 0.119

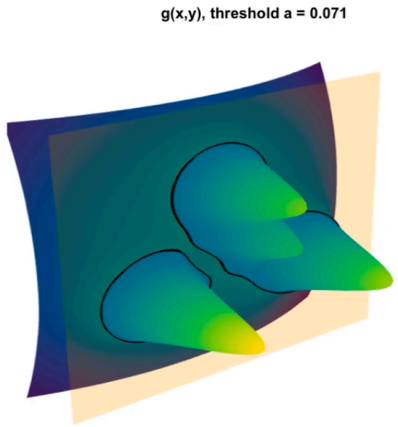

Superlevel-Set Filtration
x
y

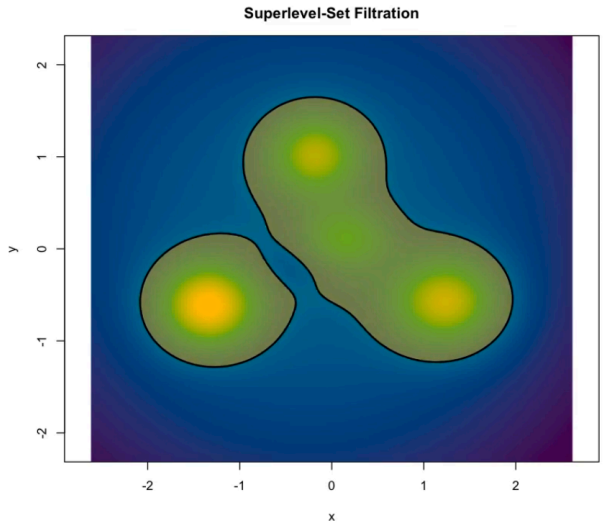

H0 Superlevel Persistence
death
birth
global max

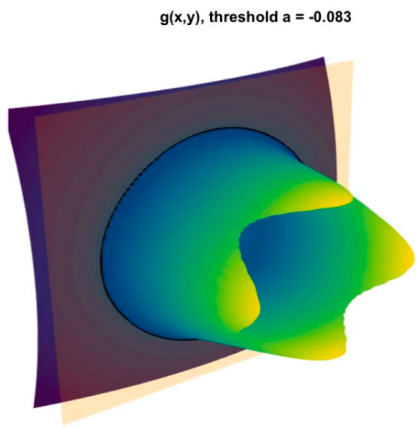

Persistence Tree — lifetime, right-to-left view

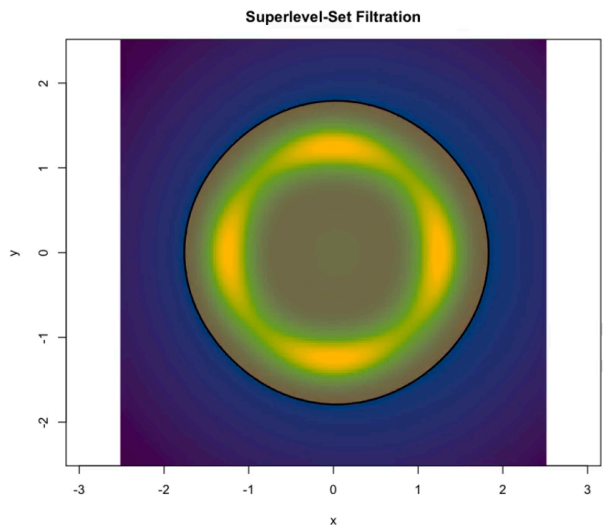

H0 and H1 Superlevel Persistence
death
birth
global max
H0
H1

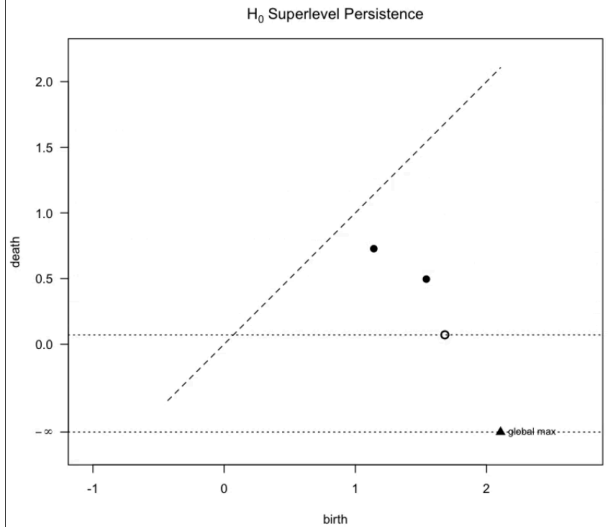

Persistence Tree — lifetime, horizontally compressed view

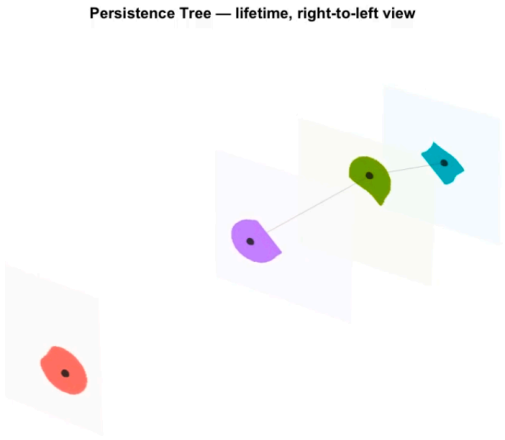

g(x,y), threshold a = 0.071

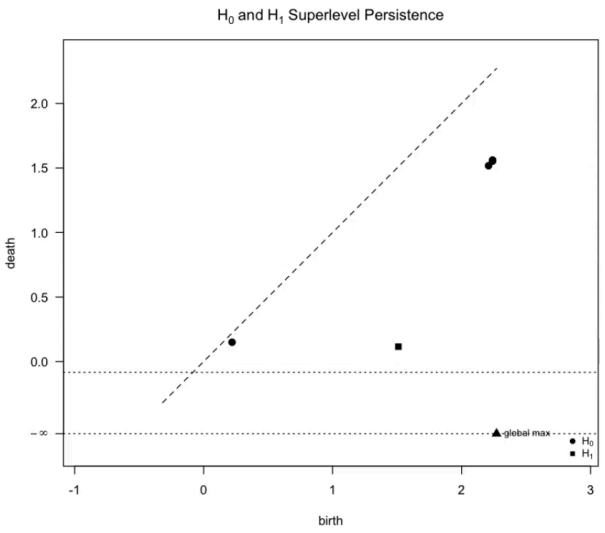

Superlevel-Set Filtration
x
y

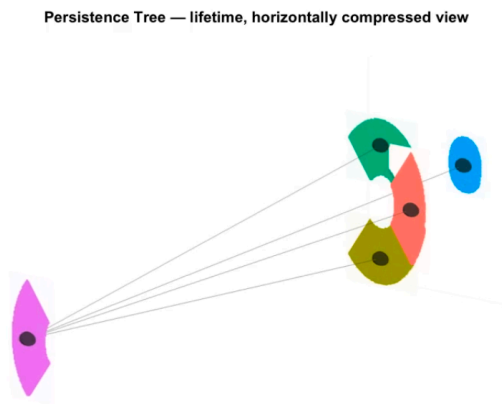

g(x,y), threshold a = -0.083

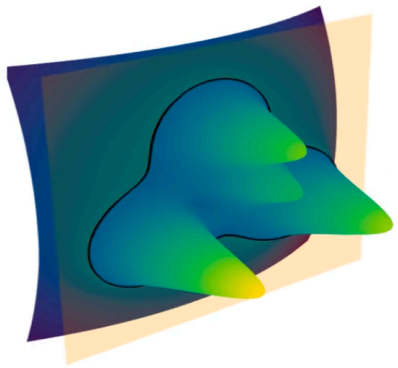

Superlevel-Set Filtration
x
y

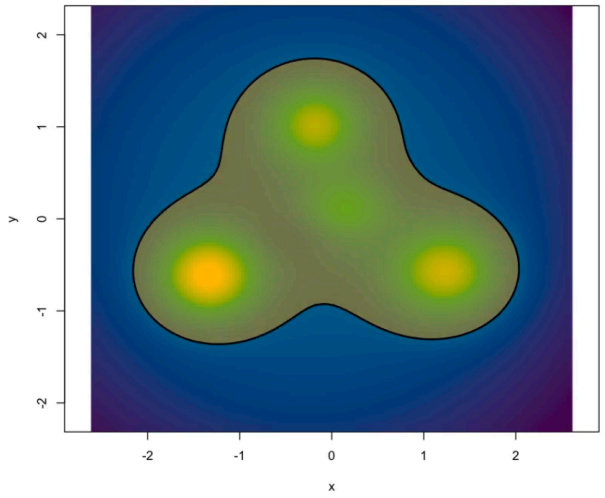

H0 Superlevel Persistence
death
birth
global max

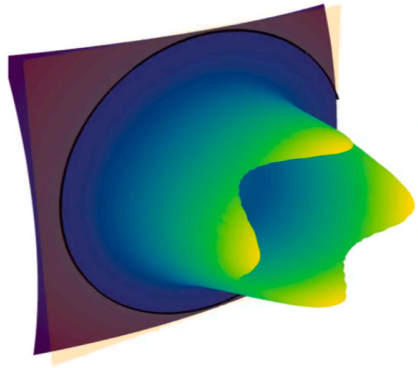

Persistence Tree — lifetime, right-to-left view

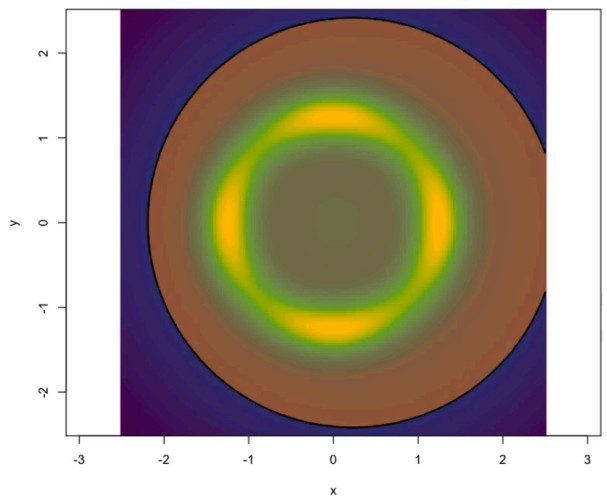

H0 and H1 Superlevel Persistence
death
birth
global max
H0
H1

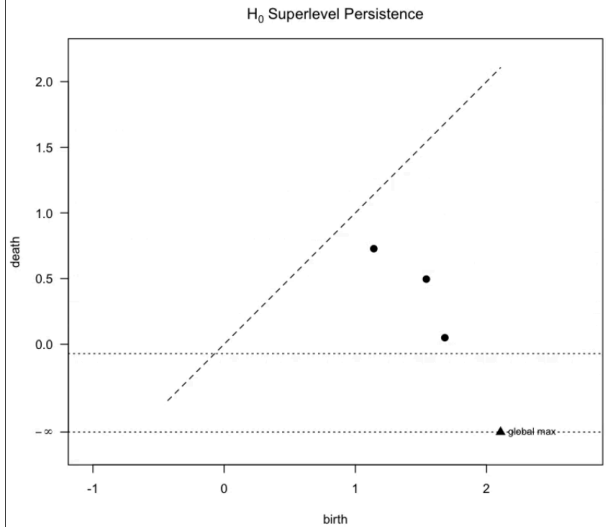

Persistence Tree — lifetime, horizontally compressed view

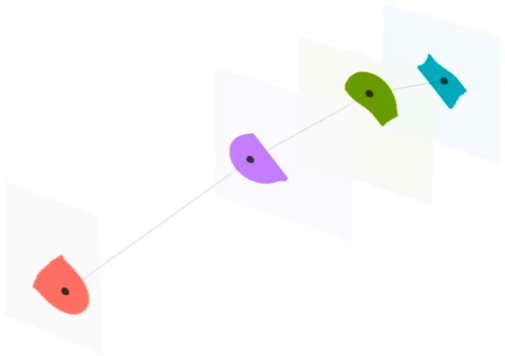

g(x,y), threshold a = -0.072

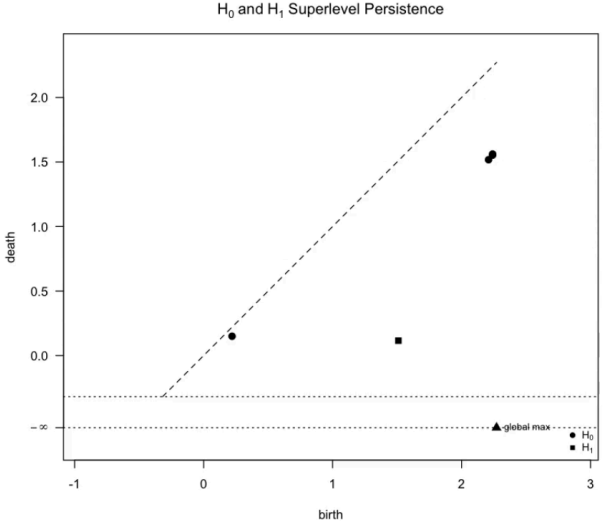

Superlevel-Set Filtration
x
y

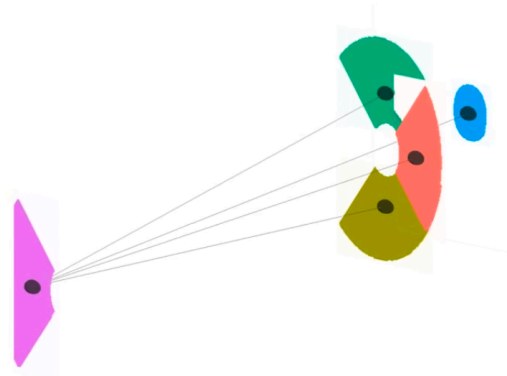

g(x,y), threshold a = -0.319


Superlevel-Set Filtration
x
y


H0 Superlevel Persistence
death
birth
global max

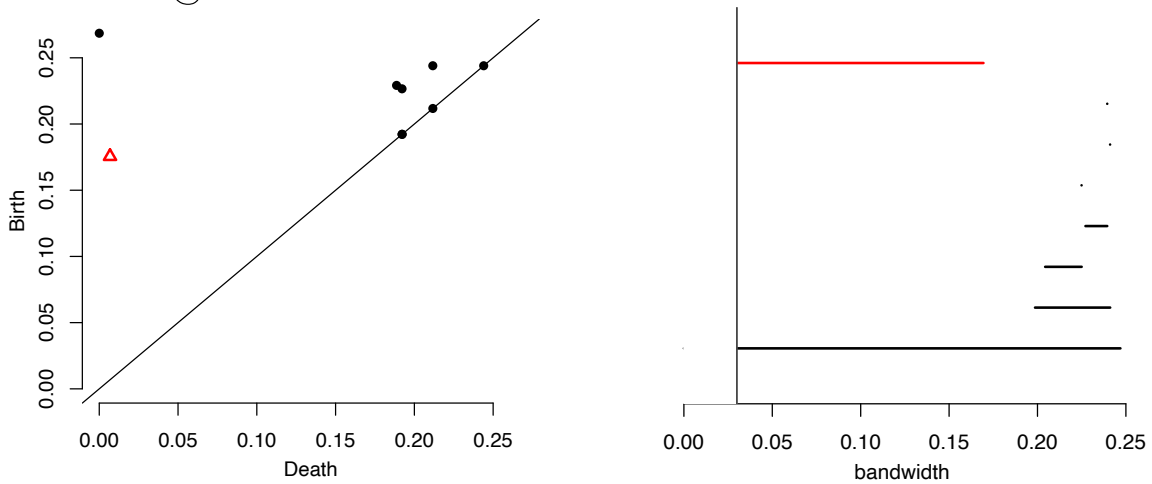

Persistence Tree — lifetime, right-to-left view


H0 and H1 Superlevel Persistence
death
birth
global max
H0
H1


Persistence Tree — lifetime, horizontally compressed view

Persistence Diagrams

from points to bars to points again

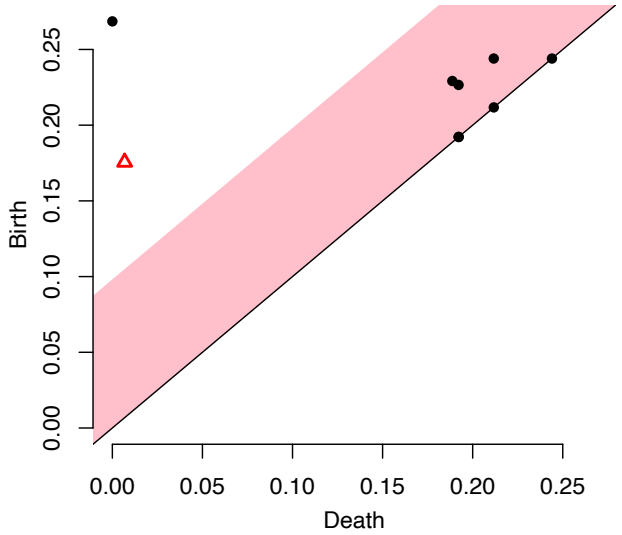


A. is interpretable:
the longer a feature appears, the more it is important (persistence)

B. is meaningful:
similar functions will have similar diagrams (stability)

Inference on the Persistence Diagrams

one word: tough



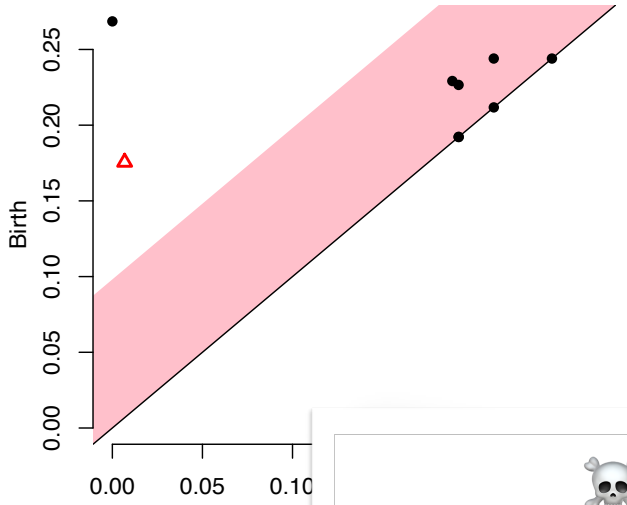
To distinguish **topological noise** from **signal** it is possible to build confidence bands on the diagonal, using:

- stability
- bootstrap

Source: Fasy, B.T., et al. (2014). Confidence sets for persistence diagrams. Ann. Statist. 42 (6) 2301-2339

Inference on the Persistence Diagrams

one word: tough



To distinguish **topological noise** from **signal** it is possible to build confidence bands on the diagonal, using:

- **stability**

☠ Achtung! Achtung! ☠
Inference using diagrams directly is not easy: they do not even have a unique mean!

Source: Turner, K., et al. (2014). Fréchet means for distributions of persistence diagrams. Discrete & Computational Geometry, 52(1)

Finanziato dall'Unione europea
NextGenerationEU

Ministero dell'Università della Ricerca

Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA

INAF
ISTITUTO NAZIONALE DI ASTROFISICA



logical noise
ssible to build
on the diagonal,

ms. Discrete & Computational Geometry, 52(1)

PNRR
ente 2
o 3.1

STILES – IR0000034
CUP C33C22000640006

Finanziato dall'Unione europea
NextGenerationEU

Ministero dell'Università e della Ricerca

Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA

INAF
ISTITUTO NAZIONALE DI ASTROFISICA



Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

Mon. Not. R. Astron. Soc. **414,** 350–383 (2011) doi:10.1111/j.1365-2966.2011.18394.x

The persistent cosmic web and its filamentary structure – I. Theory and implementation

T. Sousbie[1,2⋆]

[1]*Department of Physics, The University of Tokyo, Tokyo 113-0033, Japan*
[2]*Institut d'astrophysique de Paris & UPMC (UMR 7095), 98, bis boulevard Arago 75 014, Paris*

Accepted 2011 January 19. Received 2011 January 14; in original form 2010 September 20



Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

Mon. Not. R. Astron. Soc. **414,** 384–403 (2011) doi:10.1111/j.1365-2966.2011.18395.x

The persistent cosmic web and its filamentary structure – II. Illustrations

T. Sousbie,[1,2⋆] C. Pichon[2,3] and H. Kawahara[4]

[1]*Department of Physics, The University of Tokyo, Tokyo 113-0033, Japan*
[2]*Institut d'astrophysique de Paris & UPMC (UMR 7095), 98, bis boulevard Arago, 75 014 Paris, France*
[3]*Oxford Astrophysics, Department of Physics, Denys Wilkinson Building, Keble Road, Oxford OX1 3RH*
[4]*Department of Physics, Tokyo Metropolitan University, Hachioji, Tokyo 192-0397, Japan*

Accepted 2011 January 19. Received 2011 January 14; in original form 2010 September 21



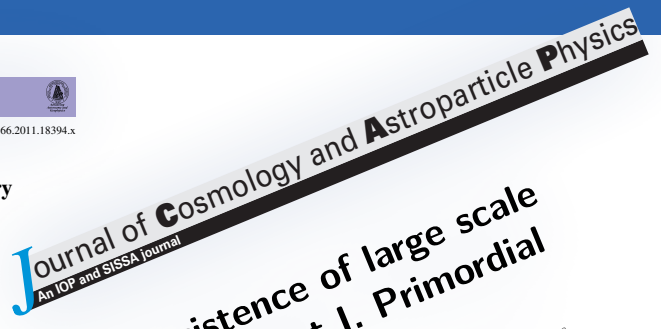
Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **507,** 2968–2990 (2021) https://doi.org/10.1093/mnras/stab232
Advance Access publication 2021 August 17

Persistent homology of the cosmic web – I. Hierarchical topology in ΛCDM cosmologies

Georg Wilding [1,2,3⋆] Keimpe Nevenzeel,[1] Rien van de Weygaert,[1,3] Gert Vegter,[2,3] Pratyush Pranav [1,4,5] Bernard J. T. Jones [1] Konstantinos Efstathiou[2,3,6] and Job Feldbrugge[7,8]

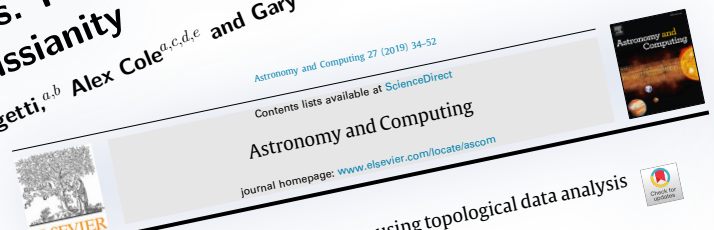
[1]*Kapteyn Astronomical Institute, University of Groningen, PO Box 800, NL-9700 AV Groningen, the Netherlands*
[2]*Bernoulli Institute for Mathematics, Computer Science and Artificial Intelligence, University of Groningen, PO Box 800, NL-9700 AV Groningen, the Netherlands*
[3]*Centre for Data Science and Systems Complexity, University of Groningen, PO Box 800, NL-9700 AV Groningen, the Netherlands*
[4]*Univ Lyon, ENS de Lyon, CNRS, Centre de Recherche Astrophysique de Lyon UMR5574, F-69007 Lyon, France*
[5]*Technion – Israel Institute of Technology, Haifa 32000, Israel*
[6]*Division of Natural and Applied Sciences and Zu Chongzhi Center for Mathematics and Computational Science, Duke Kunshan University, No. 8 Duke Avenue, Kunshan 215316, Jiangsu Province, China*
[7]*Perimeter Institute, 31 Caroline St N, Waterloo, ON N2L 2Y5, Canada*
[8]*Department of Physics, Carnegie Mellon University, 5000 Forbes Ave, Pittsburgh, PA 15217, USA*



Journal of Cosmology and Astroparticle Physics
An IOP and SISSA journal

The persistence of large scale structures. Part I. Primordial non-Gaussianity

Matteo Biagetti,[a,b] Alex Cole[a,c,d,e] and Gary Shiu[e]



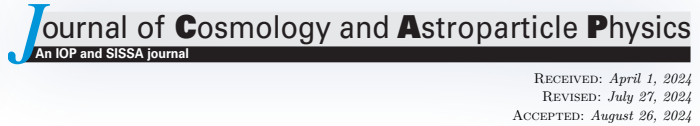
Astronomy and Computing 27 (2019) 34–52
Contents lists available at ScienceDirect
Astronomy and Computing
journal homepage: www.elsevier.com/locate/ascom

Full length article

Finding cosmic voids and filament loops using topological data analysis

X. Xu[a], J. Cisewski-Kehe[a,*], S.B. Green[b], D. Nagai[b,c,d]

[a] Department of Statistics and Data Science, Yale University, New Haven, CT 06520, USA
[b] Department of Physics, Yale University, New Haven, CT 06520, USA
[c] Department of Astronomy, Yale University, New Haven, CT 06520, USA
[d] Yale Center for Astronomy and Astrophysics, Yale University, New Haven, CT 06520, USA

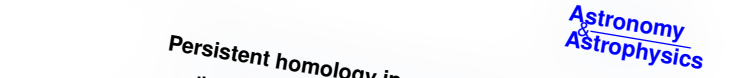


Journal of Cosmology and Astroparticle Physics
An IOP and SISSA journal

Received: *April 1, 2024*
Revised: *July 27, 2024*
Accepted: *August 26, 2024*
Published: *September 18, 2024*

Cosmology with persistent homology: a Fisher forecast

Jacky H.T. Yip [a] Matteo Biagetti [b,c,d] Alex Cole [e,f,g] Karthik Viswanathan [f,g] and Gary Shiu [a]



A&A 667, A125 (2022)
https://doi.org/10.1051/0004-6361/202243868
© S. Heydenreich et al. 2022

Astronomy & Astrophysics

Persistent homology in cosmic shear

II. A tomographic analysis of DES-Y1

Sven Heydenreich[1], Benjamin Brück[2], Pierre Burger[1], Joachim Harnois-Déraps[3,4], Sandra Unruh[1,5], Tiago Castro[6,7,8], Klaus Dolag[9,10], and Nicolas Martinet[11]



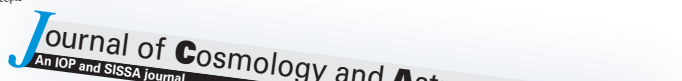
RAS Techniques and Instruments https://doi.org/10.1093/rasti/rzaf006

RASTI **4,** 1–18 (2025)
Advance Access publication 2025 February 26

DRUID: source detection and deblending in astronomical images with persistent homology

R. A. Shaw ⋆, S. Fotopoulou, M. Birkinshaw, N. Maddox and H. Stewart
School of Physics, HH Wills Physics Laboratory, Tyndall Ave, Bristol BS8 1TL, UK

Accepted 2025 February 19. Received 2025 February 17; in original form 2024 March 27



Journal of Cosmology and Astroparticle Physics
An IOP and SISSA journal

Received: *February 12, 2025*
Revised: *July 25, 2025*
Accepted: *August 14, 2025*
Published: *September 19, 2025*

Cosmology with persistent homology: parameter inference via machine learning

Juan Calles [a,1] Jacky H.T. Yip [b,1] Gabriella Contardo [c,d] Jorge Noreña [e] Adam Rouhiainen [b] and Gary Shiu [b]

PNRR
Missione 4 • Componente 2
Investimento 3.1

STILES – IR0000034
CUP C33C22000640006

Finanziato dall'Unione europea
NextGenerationEU

Ministero dell'Università e della Ricerca

Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA

INAF
ISTITUTO NAZIONALE DI ASTROFISICA



Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

Mon. Not. R. Astron. Soc. **414,** 350–383 (2011)

The persistent cosmic web and implementation

T. Sousbie[1,2]★

[1]Department of Physics, The University of Tokyo, Tokyo 113-0
[2]Institut d'astrophysique de Paris & UPMC (UMR 7095), 98,

Accepted 2011 January 19. Received 2011 January 14; in original

Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

Mon. Not. R. Astron. Soc. **414,** 384–403 (2011)

The persistent cosmic web and its filamentary stru

T. Sousbie,[1,2]★ C. Pichon[2,3] and H. Kawahara[4]

[1]Department of Physics, The University of Tokyo, Tokyo 113-0033, Japan
[2]Institut d'astrophysique de Paris & UPMC (UMR 7095), 98, bis boulevard Arago, 75 014 Paris, France
[3]Oxford Astrophysics, Department of Physics, Denys Wilkinson Building, Keble Road, Oxford OX1 3RH
[4]Department of Physics, Tokyo Metropolitan University, Hachioji, Tokyo 192-0397, Japan

Accepted 2011 January 19. Received 2011 January 14; in original form 2010 September 21

Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **507,** 2968–2990 (2021)
Advance Access publication 2021 August 17

Persistent homology of the cosmic web – I. H ΛCDM cosmologies

Georg Wilding [1,2,3]★ Keimpe Nevenzeel,[1] Rien van de Weygaert,
Pratyush Pranav [1,4,5] Bernard J. T. Jones [1] Konstantinos Efstathi

[1]Kapteyn Astronomical Institute, University of Groningen, PO Box 800, NL-9700 AV Groningen, the Ne
[2]Bernoulli Institute for Mathematics, Computer Science and Artificial Intelligence, University of Groning
Netherlands
[3]Centre for Data Science and Systems Complexity, University of Groningen, PO Box 800, NL-9700 AV Grom
[4]Univ Lyon, ENS de Lyon, CNRS, Centre de Recherche Astrophysique de Lyon UMR5574, F-69007 Lyon, Fra
[5]Technion – Israel Institute of Technology, Haifa 32000, Israel
[6]Division of Natural and Applied Sciences and Zu Chongzhi Center for Mathematics and Computational Scienc
Avenue, Kunshan 215316, Jiangsu Province, China
[7]Perimeter Institute, 31 Caroline St N, Waterloo, ON N2L 2Y5, Canada
[8]Department of Physics, Carnegie Mellon University, 5000 Forbes Ave, Pittsburgh, PA 15217, USA

Topological persistence for astronomical image segmentation

Segmentazione di immagini astronomiche mediante strumenti topologici

Ceccaroni, R., Brutti, P., Castellano, M., Fontana, A. and Merlin, E.

Abstract The fast evolution of telescope technologies is makin
tion of a massive wide variety of data. In our era, a tool that
extraction from astronomical images is essential. Within thi
mentation plays a key role and classical edge-based de
well-suited to deal with astronomical images because th
boundary structure.
Thus, to effectively tackle this task, it is mandatory
The main goal of this work is to design and test a
method strongly based on Topological Data Ana
tools like persistent homology and persistenc
identify the connected components of abstra
them to use in order to compute a sensible
Abstract La rapida evoluzione delle tec
bile la raccolta di un'ampia varietà di
mento che automatizza l'estrazione d
All'interno di questo ampio compi
ruolo chiave. Gli algoritmi classici
adatti per le immagini astronom
dei confini ben definiti.

Riccardo Ceccaroni
Sapienza University of Roma
Pierpaolo Brutti
Sapienza University of P

Topological Attention for Denoising Astronomical Images

Riccardo Ceccaroni and Pierpaolo Brutti

Abstract Astronomical observations often involve capturing faint signals from ce-
lestial objects, such as distant galaxies or dim stars. These signals can be easily
overwhelmed by noise, which includes electronic noise, atmospheric interference,
and other artifacts. It is then clear that using an effective denoising algorithm is
crucial to improve the accuracy of data interpretation by extracting the actual as-
tronomical signals from the noisy background. In this work, we introduce a novel
denoising approach based on deep neural networks with pixel attention derived from
the topological structure extracted by 0-dimensional Persistent Homology (PH). Our
network architecture combines an AutoEncoder (AE) with a Topological Module
(TM) that assigns weights to pixels based on PH, providing nuanced attention to
the relevant image regions. Our methodology is evaluated against widely used de-
noising methods, showcasing its potential for enhancing the quality of astronomical
images.

Key words: Denoising, Astronomical Image, Persistent Homology, Topological Attention

1 Introduction

In the ever-expanding field of astronomical data analysis, obtaining high-qualit
images is key. However, atmospheric fluctuations, optical distortions,
calibration, and inherent noise during acquisition, may strongl

Riccardo Ceccaroni
Università di Roma "La Sapien
e-mail: riccardo.cecc
Pierp

Astronomy & Astrophysics

nruh[1,5], Tiago Castro[6,7,8],

https://doi.org/10.1093/rasti/rzaf006

stronomical images with

and H. Stewart

h 27

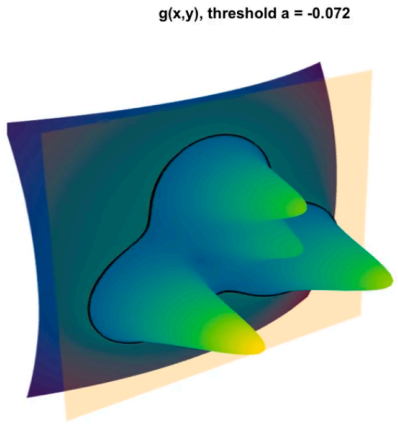
and Astroparticle Physics

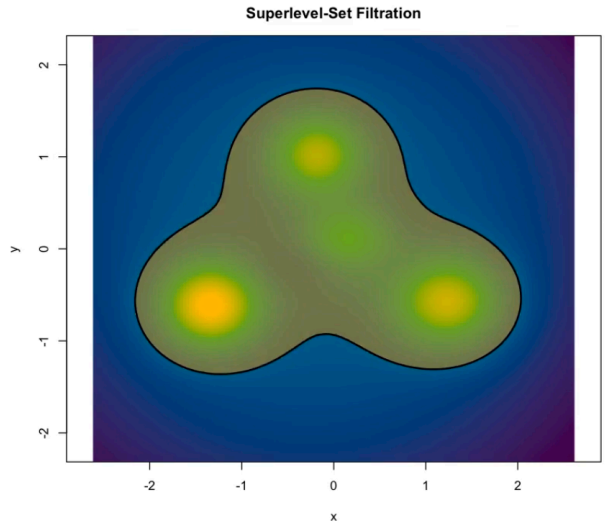
Received: February 12, 2025
Revised: July 25, 2025
Accepted: August 14, 2025
Published: September 19, 2025

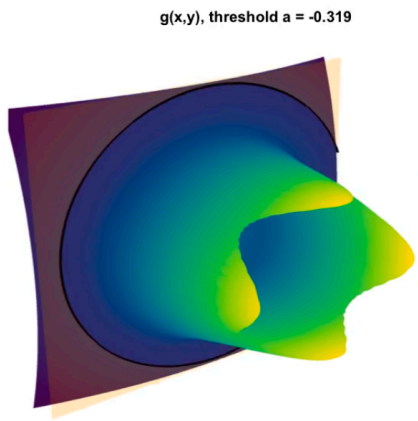
t homology: parameter arning

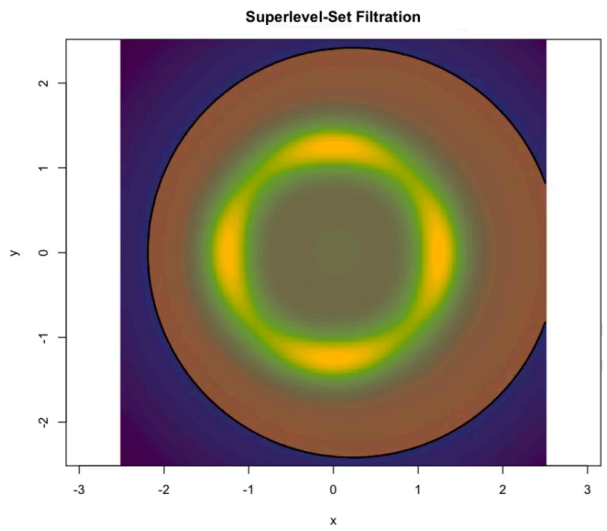
Gabriella Contardo [c,d] Jorge Noreña [e]
[b]

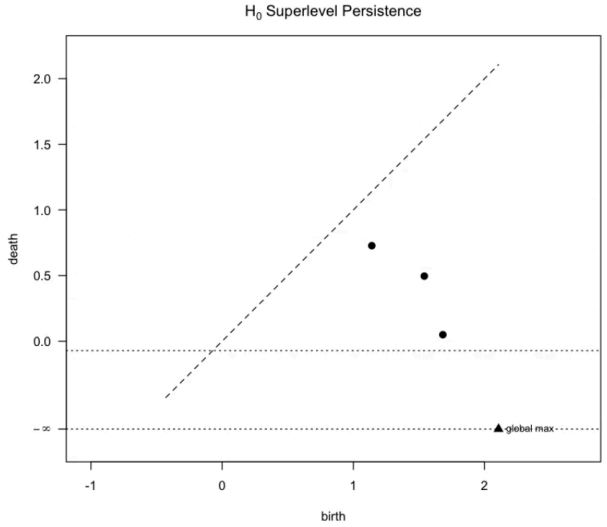
. . . nevertheless . . .

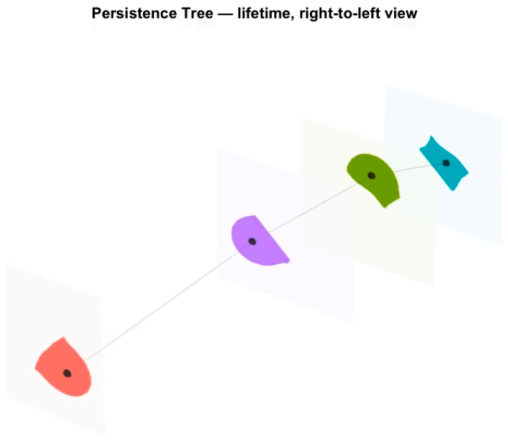

g(x,y), threshold a = -0.072

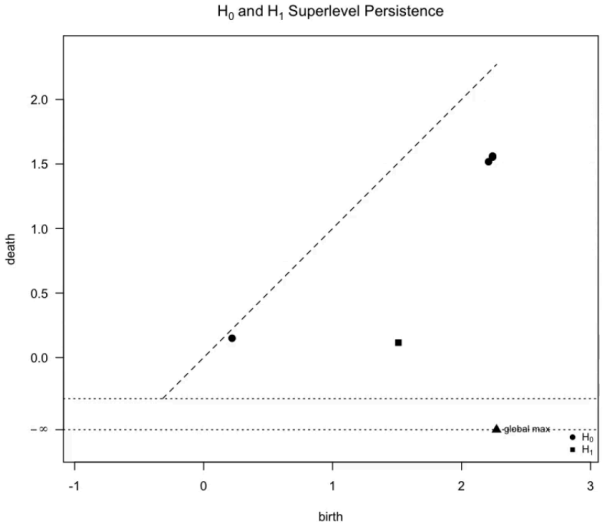

Superlevel-Set Filtration
x
y

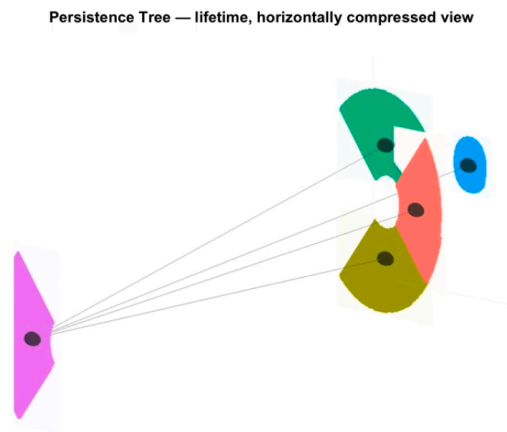

g(x,y), threshold a = -0.319


Superlevel-Set Filtration
x
y


H0 Superlevel Persistence
death
birth
global max


Persistence Tree — lifetime, right-to-left view


H0 and H1 Superlevel Persistence
death
birth
global max
H0
H1

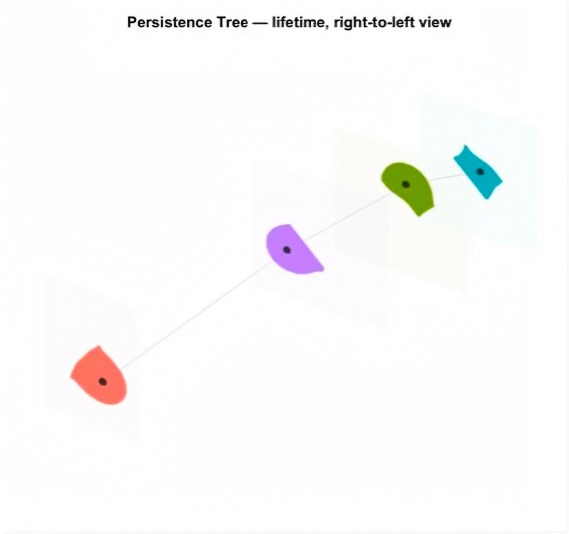

Persistence Tree — lifetime, horizontally compressed view










Persistence Tree — lifetime, right-to-left view


Persistence Tree — lifetime, horizontally compressed view

Finanziato dall'Unione europea
NextGenerationEU

Ministero dell'Università e della Ricerca

Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA

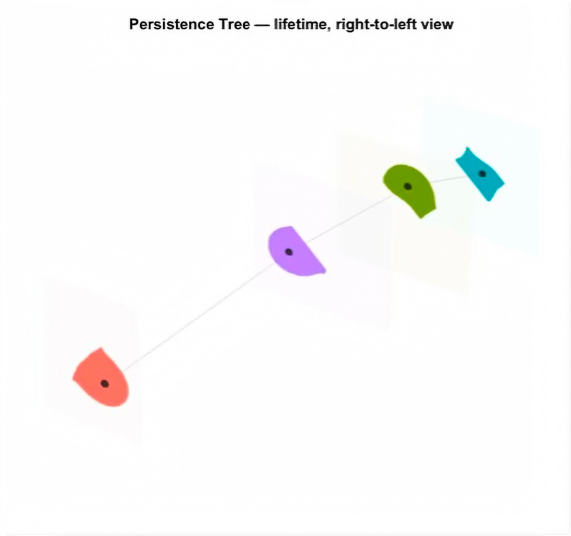
INAF
ISTITUTO NAZIONALE DI ASTROFISICA

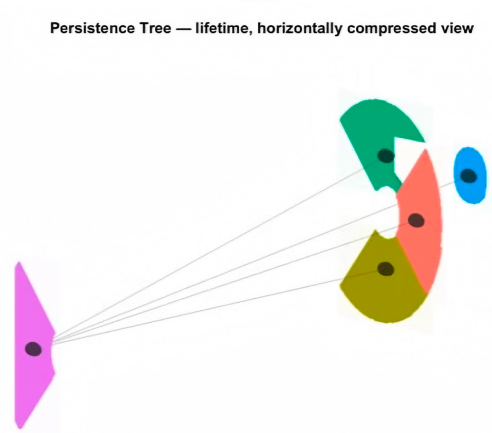














Finanziato dall'Unione europea
NextGenerationEU

Ministero dell'Università e della Ricerca

Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA

INAF
ISTITUTO NAZIONALE DI ASTROFISICA

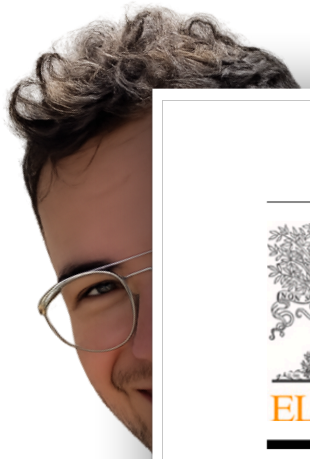


Image and Vision Computing 175 (2026) 106154

Contents lists available at ScienceDirect

Image and Vision Computing

journal homepage: www.elsevier.com/locate/imavis

Persistence Trees: A topological tree structure for anisotropic image denoising

Riccardo Ceccaroni [a,*], Pierpaolo Brutti [a], Anastasia Zakharova [b], Thierry Bouwmans [b]

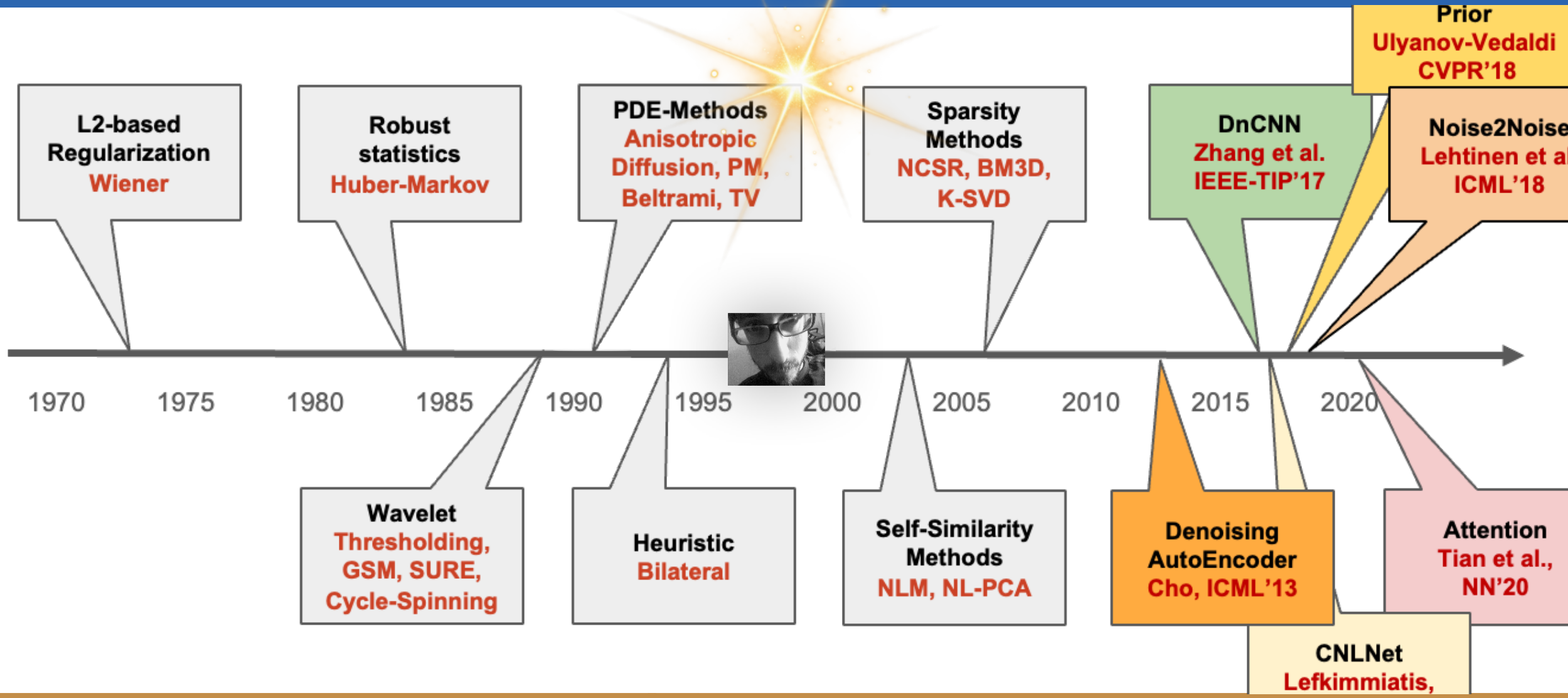
[a] Università di Roma, La Sapienza, Piazzale Aldo Moro, 5, Roma, 00185, Italy
[b] Laboratoire MIA, La Rochelle Université, Av. Henri Becquerel, La Rochelle, 17000, France

ARTICLE INFO

Keywords:
Persistent homology
Anisotropic diffusion
Topological denoising
Image restoration

ABSTRACT

Image denoising is a fundamental task in image processing, as it is often the first step in subsequent image enhancement, analysis, and interpretation. Although conventional methods are computationally efficient, they often fail to preserve fine image details when the signal-to-noise ratio is low. More recent approaches exploit nonlocal similarities or advanced regularization strategies; however, they remain largely intensity-driven and can be sensitive to geometric variability and texture complexity. To address these limitations, we introduce


L2-based Regularization Wiener
Robust statistics Huber-Markov
PDE-Methods Anisotropic Diffusion, PM, Beltrami, TV
Sparsity Methods NCSR, BM3D, K-SVD
DnCNN Zhang et al. IEEE-TIP'17
Prior Ulyanov-Vedaldi CVPR'18
Noise2Noise Lehtinen et al ICML'18
1970
1975
1980
1985
1990
1995
2000
2005
2010
2015
2020
Wavelet Thresholding, GSM, SURE, Cycle-Spinning
Heuristic Bilateral
Self-Similarity Methods NLM, NL-PCA
Denoising AutoEncoder Cho, ICML'13
CNLNet Lefkimmiatis,
Attention Tian et al., NN'20

Finanziato dall'Unione europea
NextGenerationEU

Ministero dell'Università e della Ricerca

Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA

INAF
ISTITUTO NAZIONALE DI ASTROFISICA



L2-based Regularization
Wiener

Robust statistics
Huber-Markov

PDE-Methods
Anisotropic Diffusion, PM, Beltrami, TV

Sparsity Methods
NCSR, BM3D, K-SVD

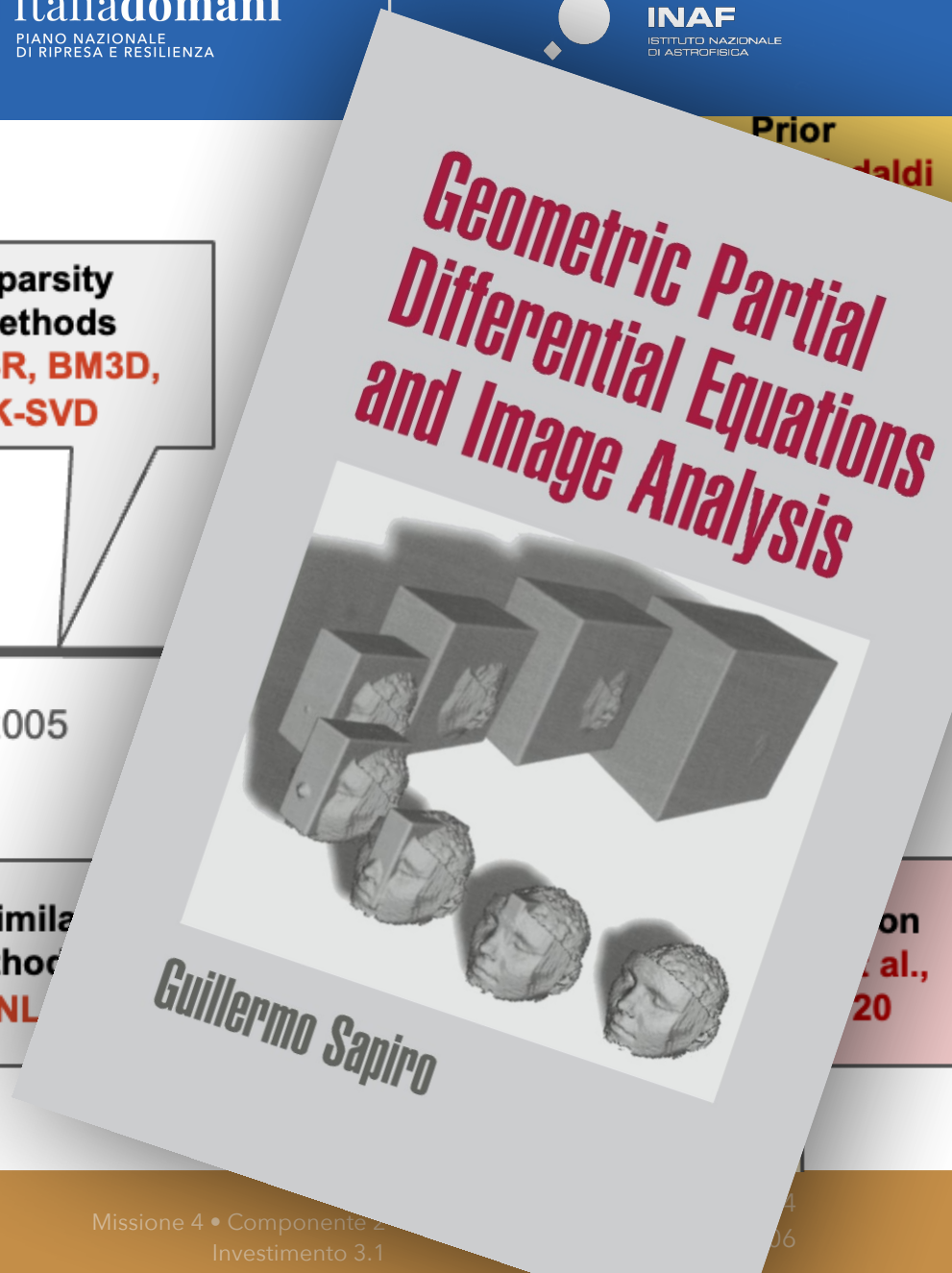


1970 1975 1980 1985 1990 1995 2000 2005

Wavelet
Thresholding, GSM, SURE, Cycle-Spinning

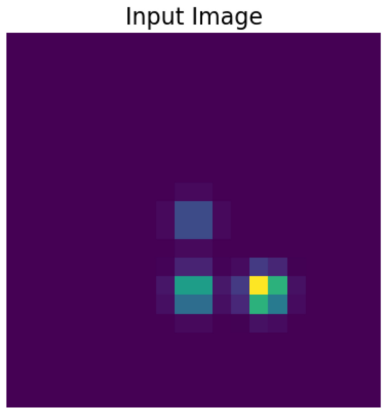
Heuristic
Bilateral

Self-Simila
Metho
NLM, NL

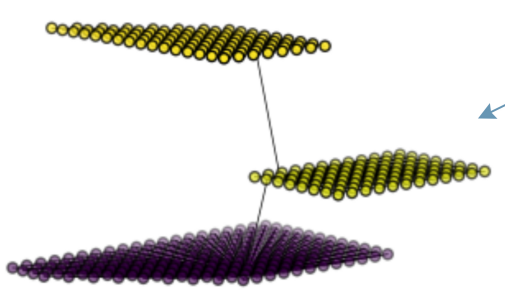


Persistence Tree for Denoising

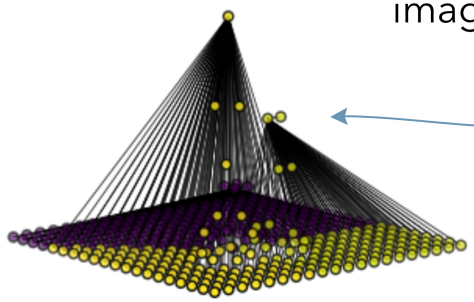
1, 10, 1000 diffusions



alternative representation: z-axis here represents persistence (**lifetime**)



spanning tree that reflects the birth/death hierarchy of **connected image components**

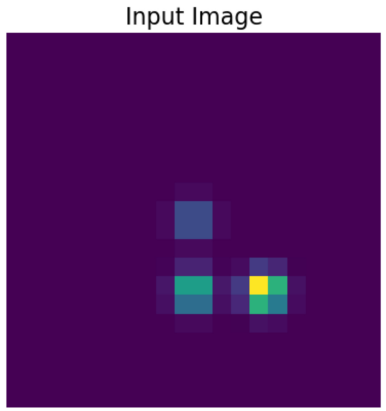


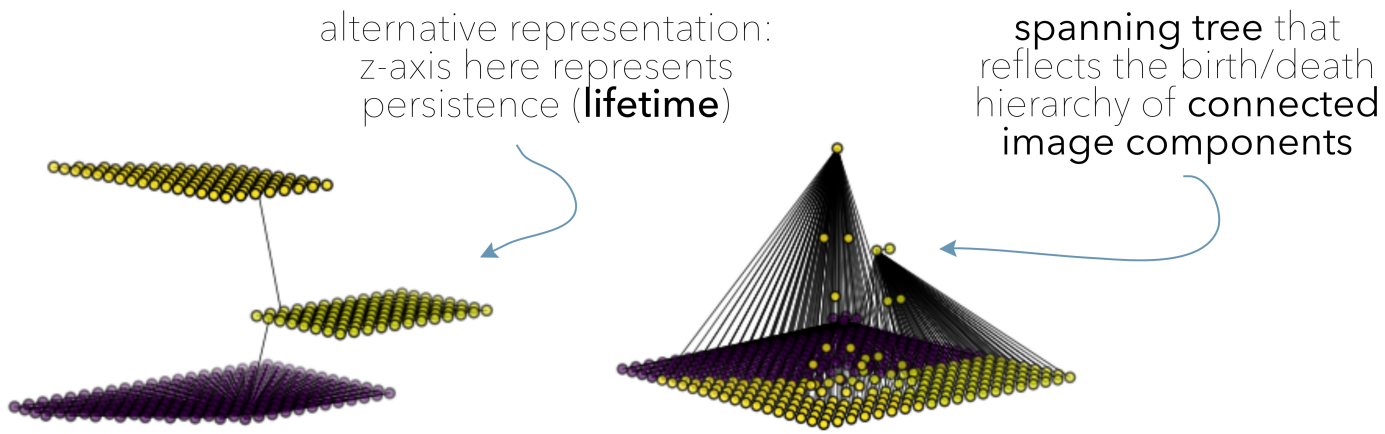
Idea

Replace the purely local, gradient-driven geometry of classical anisotropic diffusion with a **topological hierarchy** extracted from the image

Persistence Tree for Denoising

1, 10, 1000 diffusions



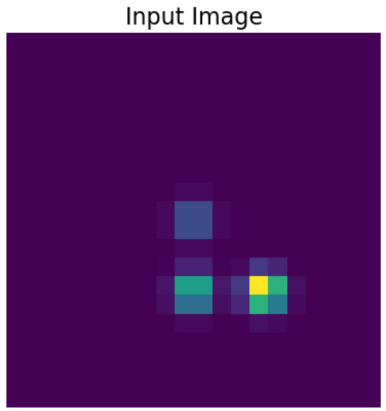


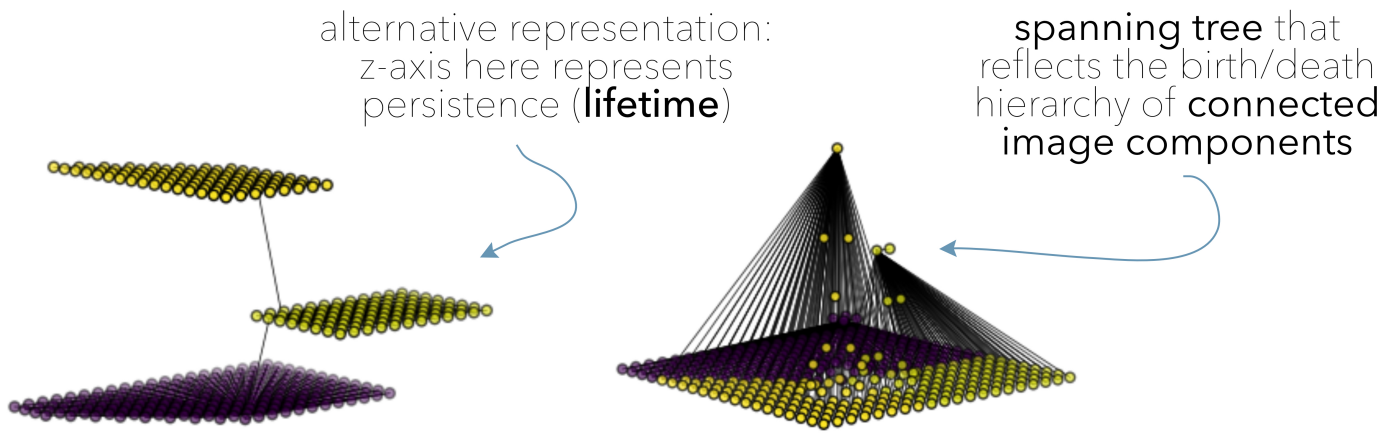
1. Perona-Malik central question:
   "Is the local gradient large enough that the diffusion should be inhibited there?"
2. Effective at preserving **strong edges**, but the gradient itself becomes unreliable under severe noise (can look like an edge)

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
INAF ISTITUTO NAZIONALE DI ASTROFISICA

Persistence Tree for Denoising

1, 10, 1000 diffusions





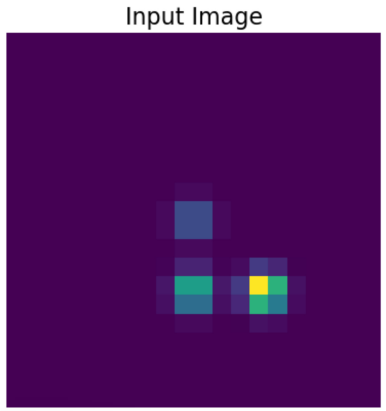
1. Perona-Malik central question:
   "Is the local gradient large enough that the diffusion should be inhibited there?"
2. Effective at preserving **strong edges**, but the gradient itself becomes unreliable under severe noise (can look like an edge)

1. Persistence Tree asks:
   "Is this intensity structure persistent enough to be considered structurally important and stop the diffusion?"
2. A feature with large persistence receives little diffusion, while a short-lived feature is smoothed more aggressively (hierarchical/multiscale vs differential/local)

PNRR Missione 4 • Componente 2 Investimento 3.1
STILES – IR0000034 CUP C33C22000640006

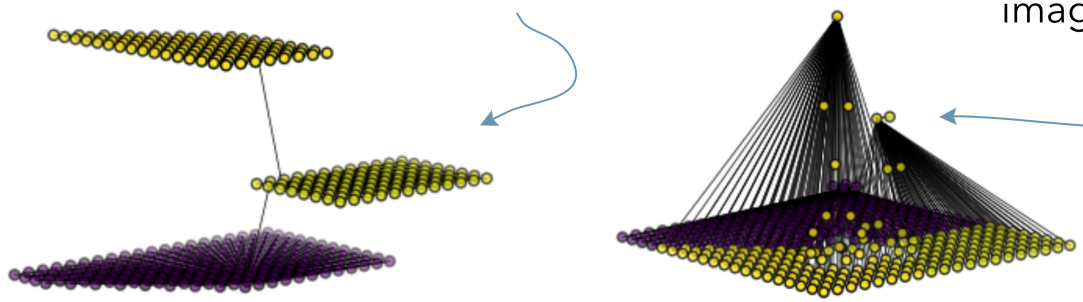
Persistence Tree for Denoising

1, 10, 1000 diffusions



alternative representation: z-axis here represents persistence (**lifetime**)

spanning tree that reflects the birth/death hierarchy of connected image components



1. Perona-Malik cer
   "Is the local grad
   should be inhibit
2. Effective at prese
   itself becomes u
   like an edge)

☠ Achtung! Achtung! ☠

Not simply Perona–Malik applied to a tree.

It is a hybrid diffusion: topological propagation along the PT + local intensity-sensitive propagation in the original image grid

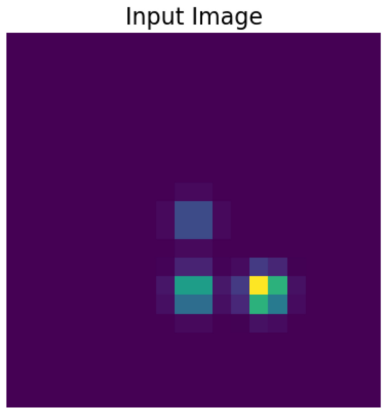
ough to be
stop the diffusion?"

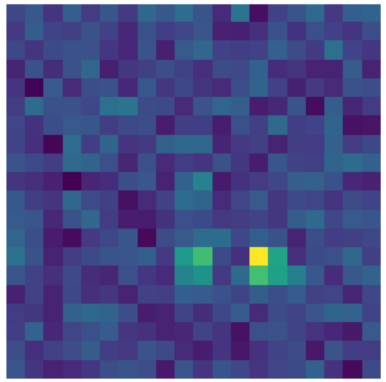
es little diffusion,
d more
aggressively (hierarchical/multiscale vs differential/local)

Persistence Tree for Denoising

robustness

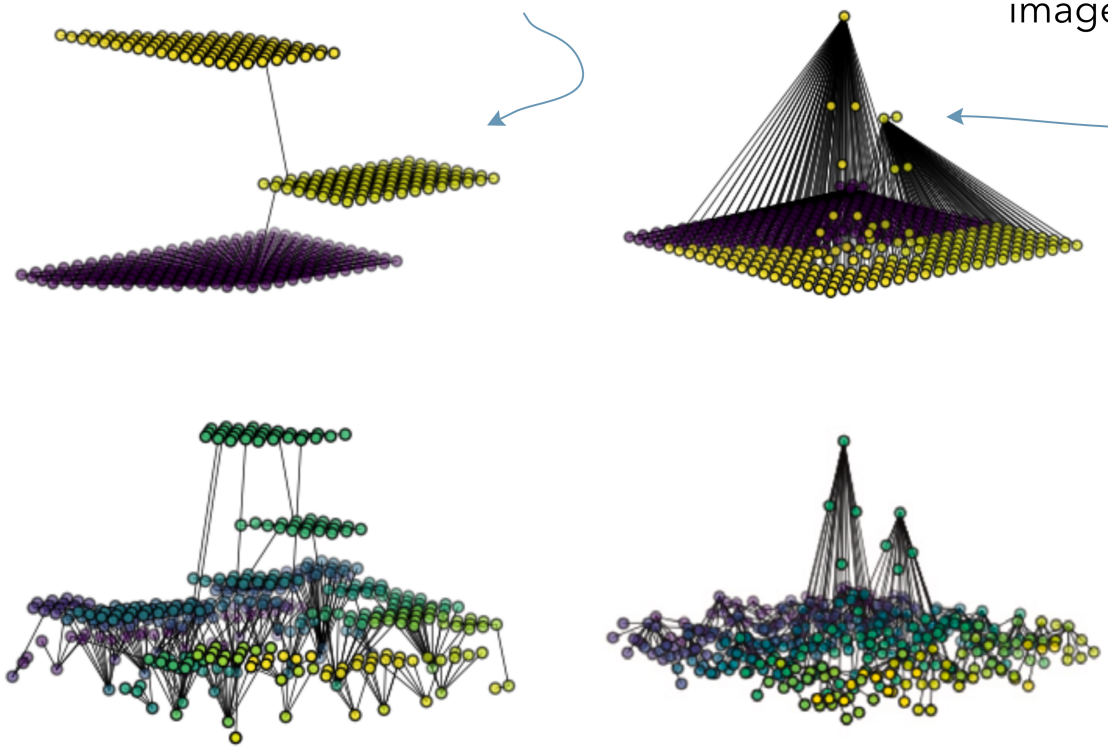
- Experimentally **filtered** and **unfiltered** trees produce almost identical denoising results despite **substantial** removal of low-persistence branches
- Theoretically **Pruning** low-persistence branches has a **theoretically bounded effect** on the resulting diffusion



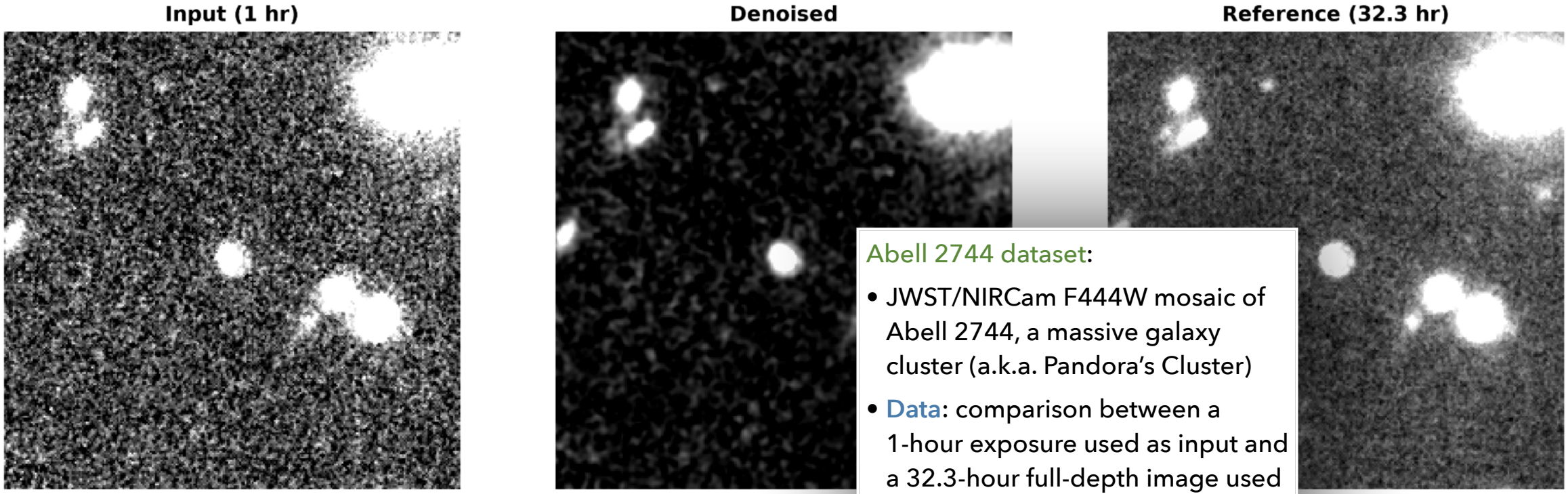


alternative representation: z-axis here represents persistence (**lifetime**)

spanning tree that reflects the birth/death hierarchy of connected image components



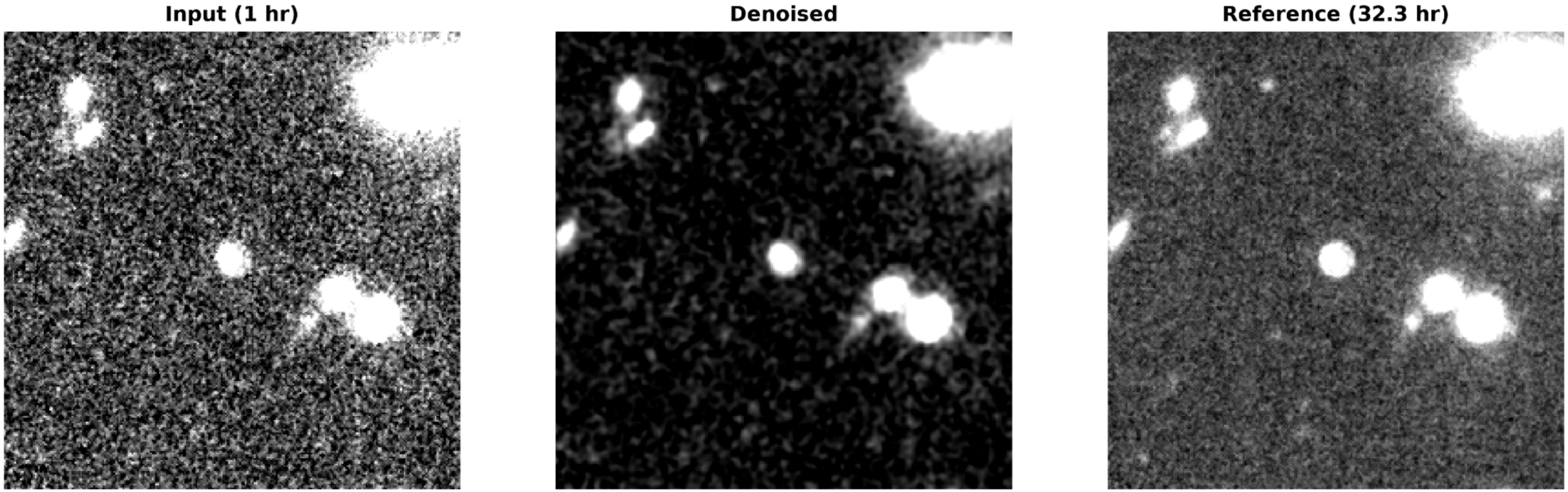
Results?



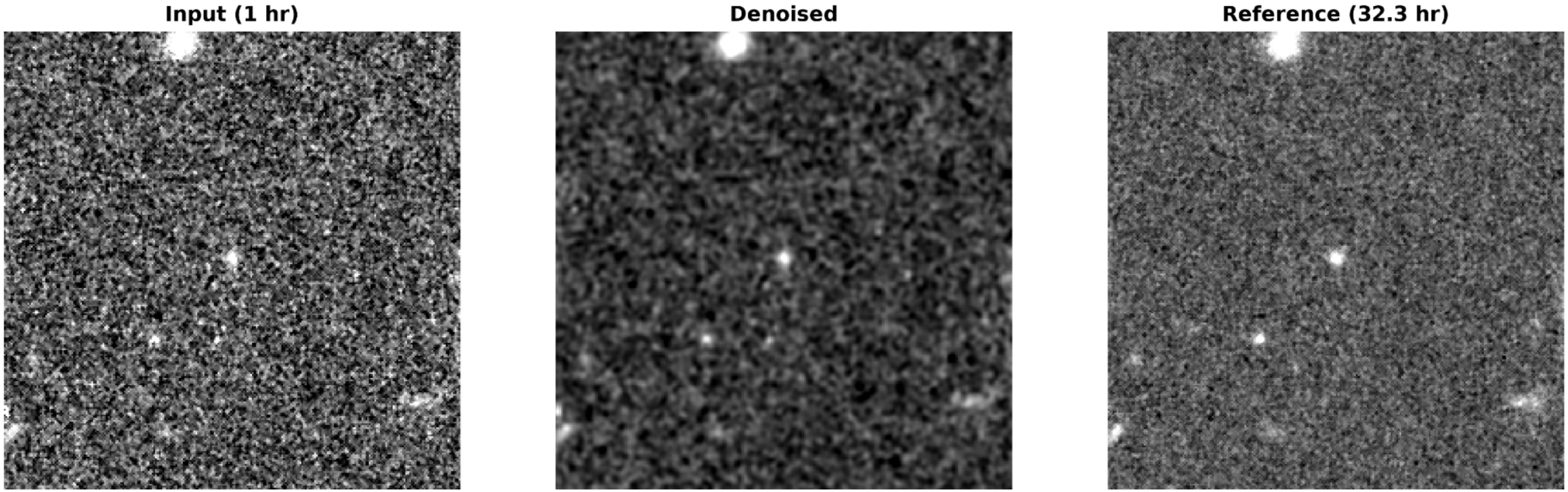
range [z1, z2] = [-8.297e-05, 2.311e-04] (estimated from the GT patch)

Abell 2744 dataset:

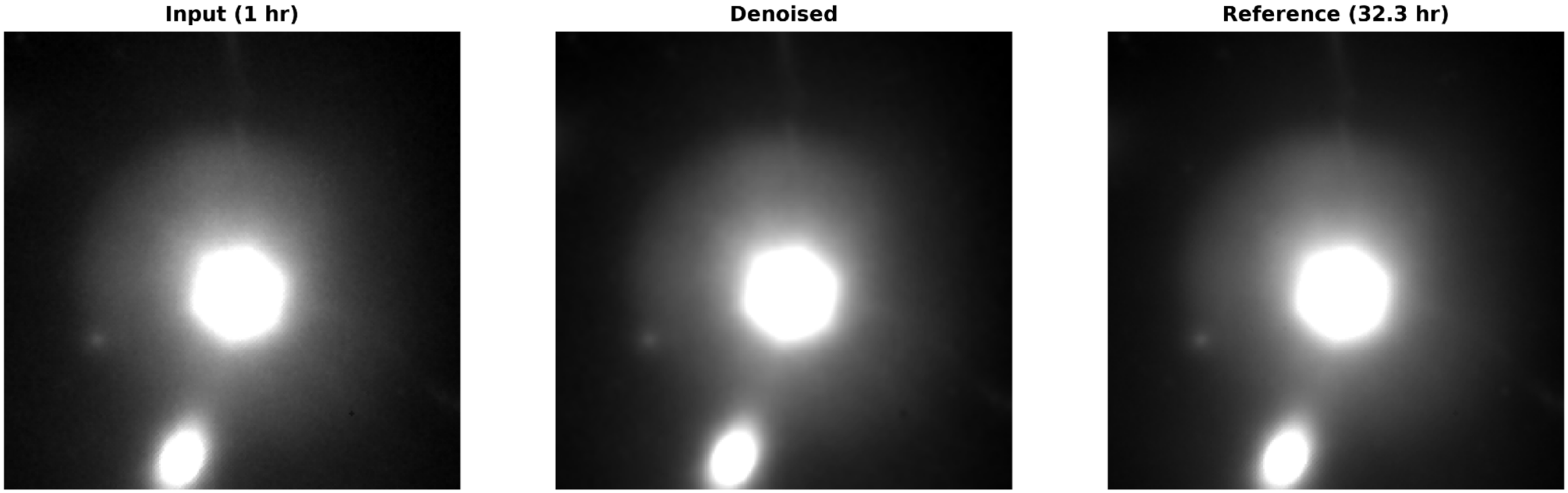
- JWST/NIRCam F444W mosaic of Abell 2744, a massive galaxy cluster (a.k.a. Pandora's Cluster)
- Data: comparison between a 1-hour exposure used as input and a 32.3-hour full-depth image used as reference.
- Frame size: 24,397 × 21,040 pixels (random patch 250 x 250)



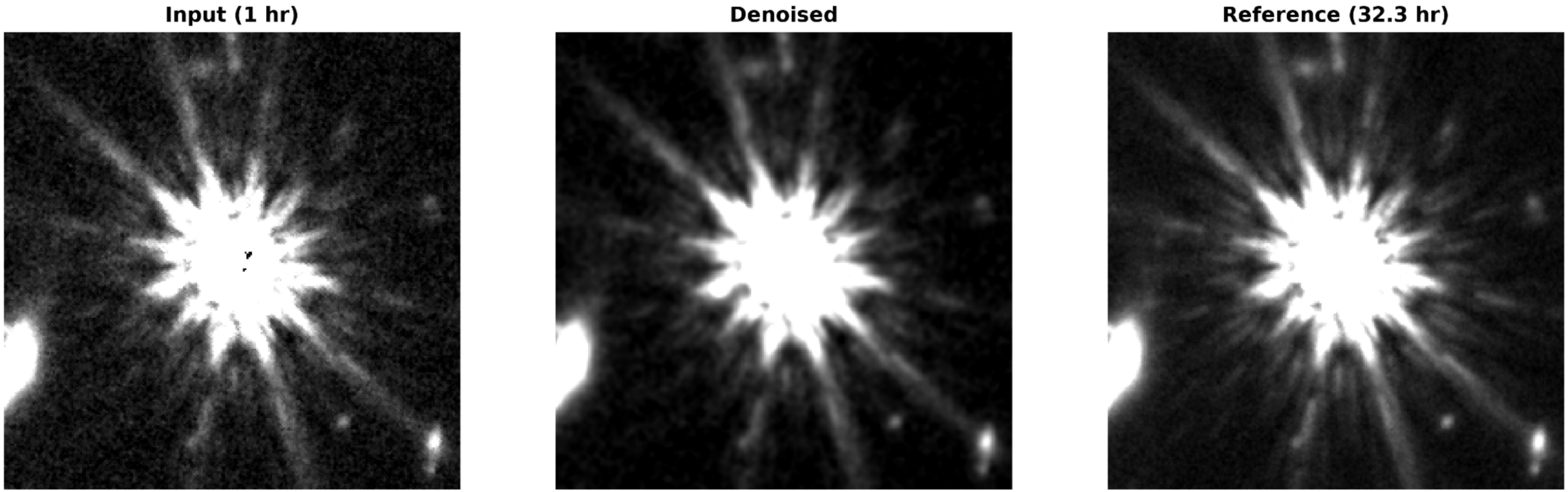
range [z1, z2] = [-8.297e-05, 2.311e-04] (estimated from the GT patch)



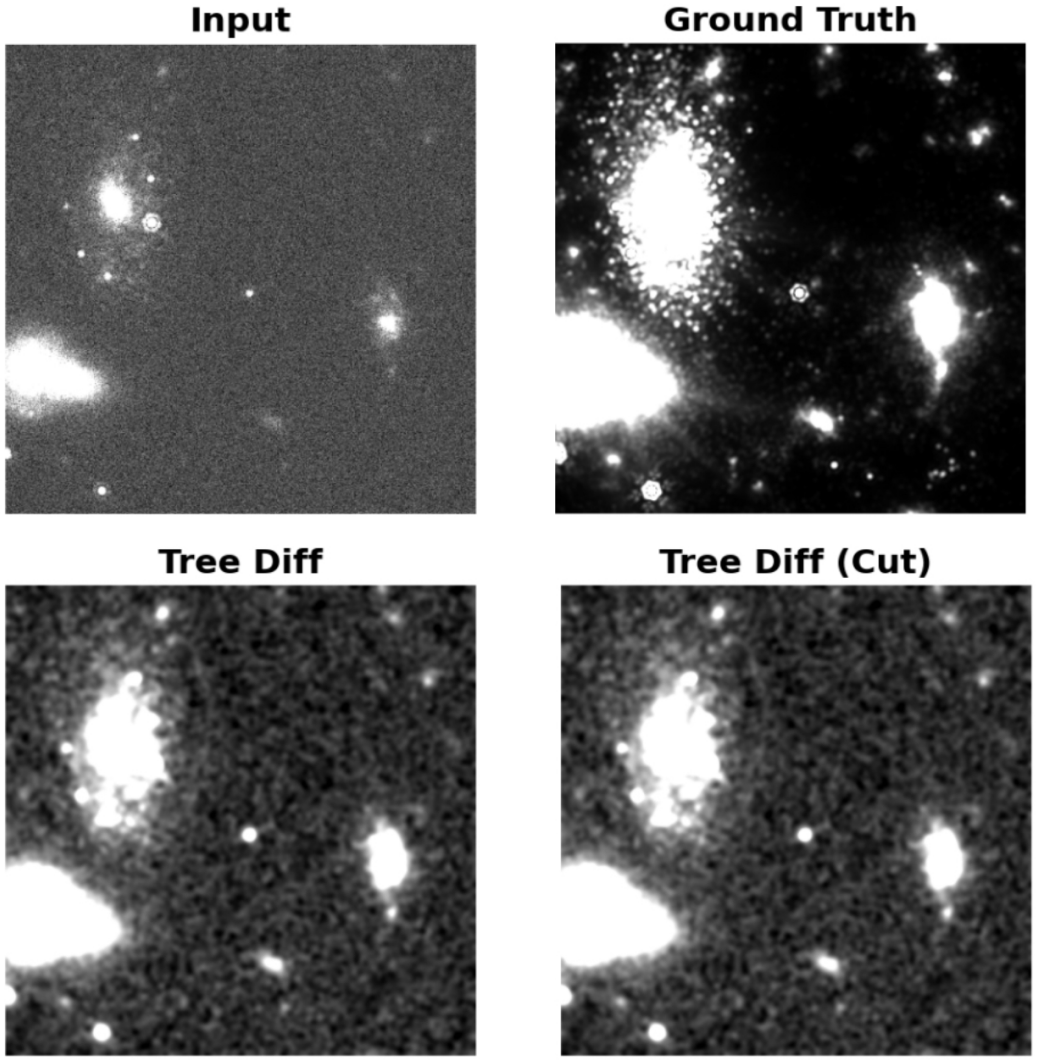
range [z1, z2] = [-2.252e-04, 4.263e-04] (estimated from the GT patch)



range [z1, z2] = [1.005e-03, 2.631e-02] (estimated from the GT patch)



range [z1, z2] = [-5.417e-05, 1.511e-03] (estimated from the GT patch)

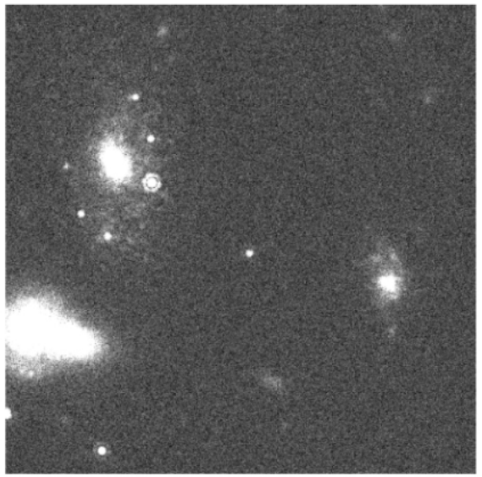


Method	MSE	PSNR (dB)
Unfiltered Tree Diffusion	**0.000029 ± 0.000068**	51.05 ± 6.49
Filtered Tree Diffusion (MaxJump)	**0.000029 ± 0.000067**	**51.07 ± 6.49**
Perona–Malik	0.002699 ± 0.011151	46.26 ± 11.77
Gaussian Filter	0.005636 ± 0.015682	31.72 ± 8.59
Median Filter	0.005638 ± 0.015690	31.72 ± 8.59
Wavelet Denoising	0.005368 ± 0.015535	31.91 ± 8.48
Non-Local Means	0.005359 ± 0.015515	31.92 ± 8.48
BM3D	0.005342 ± 0.015599	32.65 ± 9.17
U-Net (Supervised)	0.000616 ± 0.002075	39.06 ± 5.64
DnCNN (Supervised)	0.001147 ± 0.004223	39.46 ± 6.91
NAFNet (Supervised)	0.000089 ± 0.000215	44.70 ± 4.72
Restormer (Supervised)	0.000032 ± 0.000017	45.50 ± 1.99
HIR-Diff (Unsupervised)	0.000058 ± 0.000132	47.09 ± 4.83

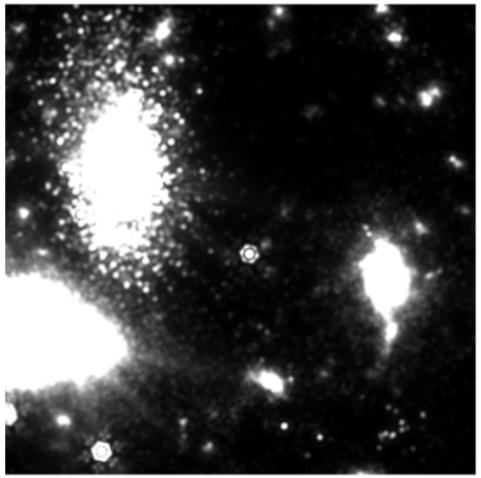
FORECAST dataset:

- Consists of high-resolution wide-field astronomical imagery capturing galaxies, nebulae, and deep sky objects.
- The original image has dimensions 26, 000 × 26, 000 pixels. Clean ground truth is available.
- To enable practical evaluation, we extract 1000 non-overlapping patches of 500 × 500 pixels.

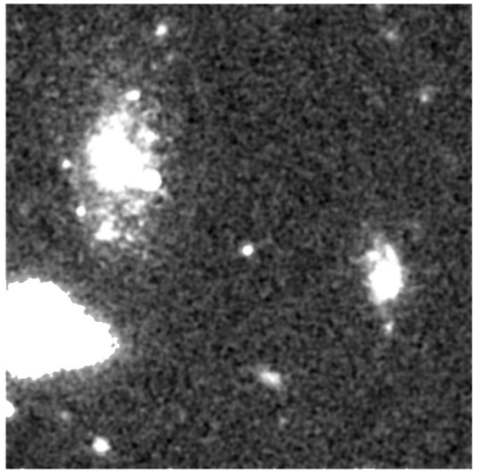
Input



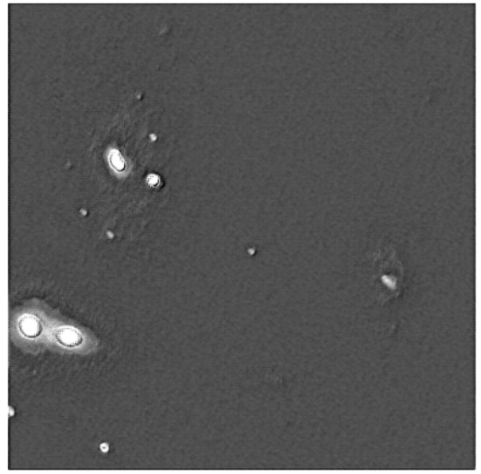
Ground Truth



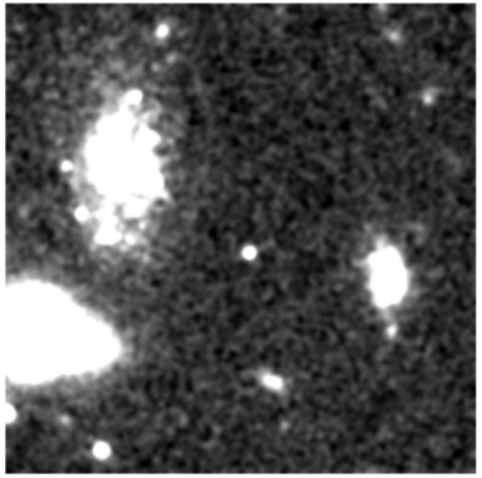
PeronaMalik



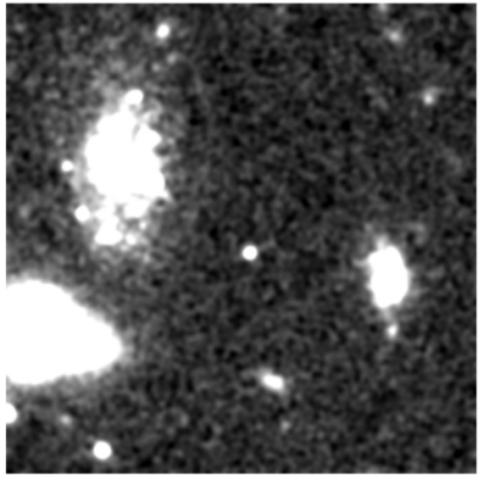
DnCNN



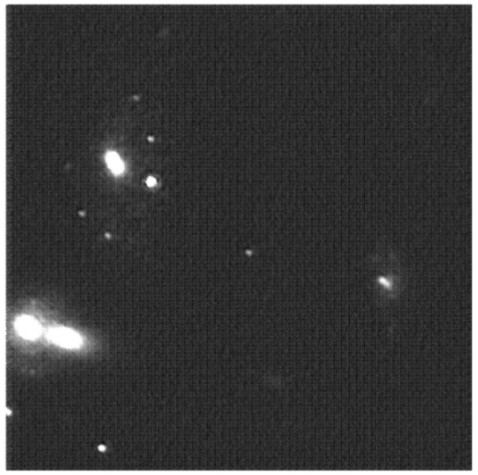
Tree Diff



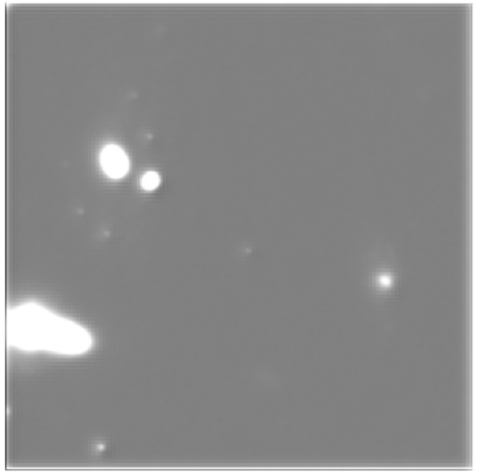
Tree Diff (Cut)

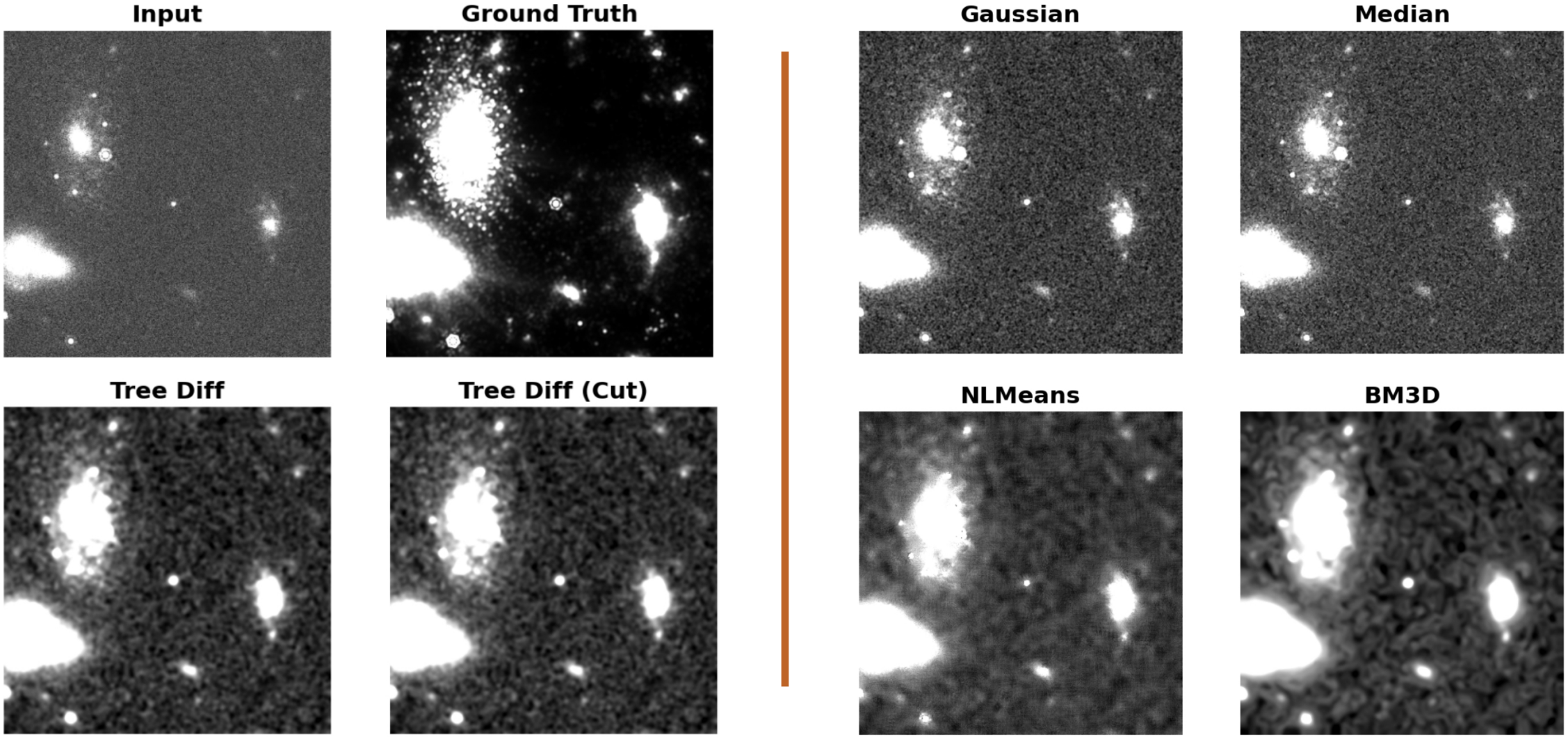


Restormer

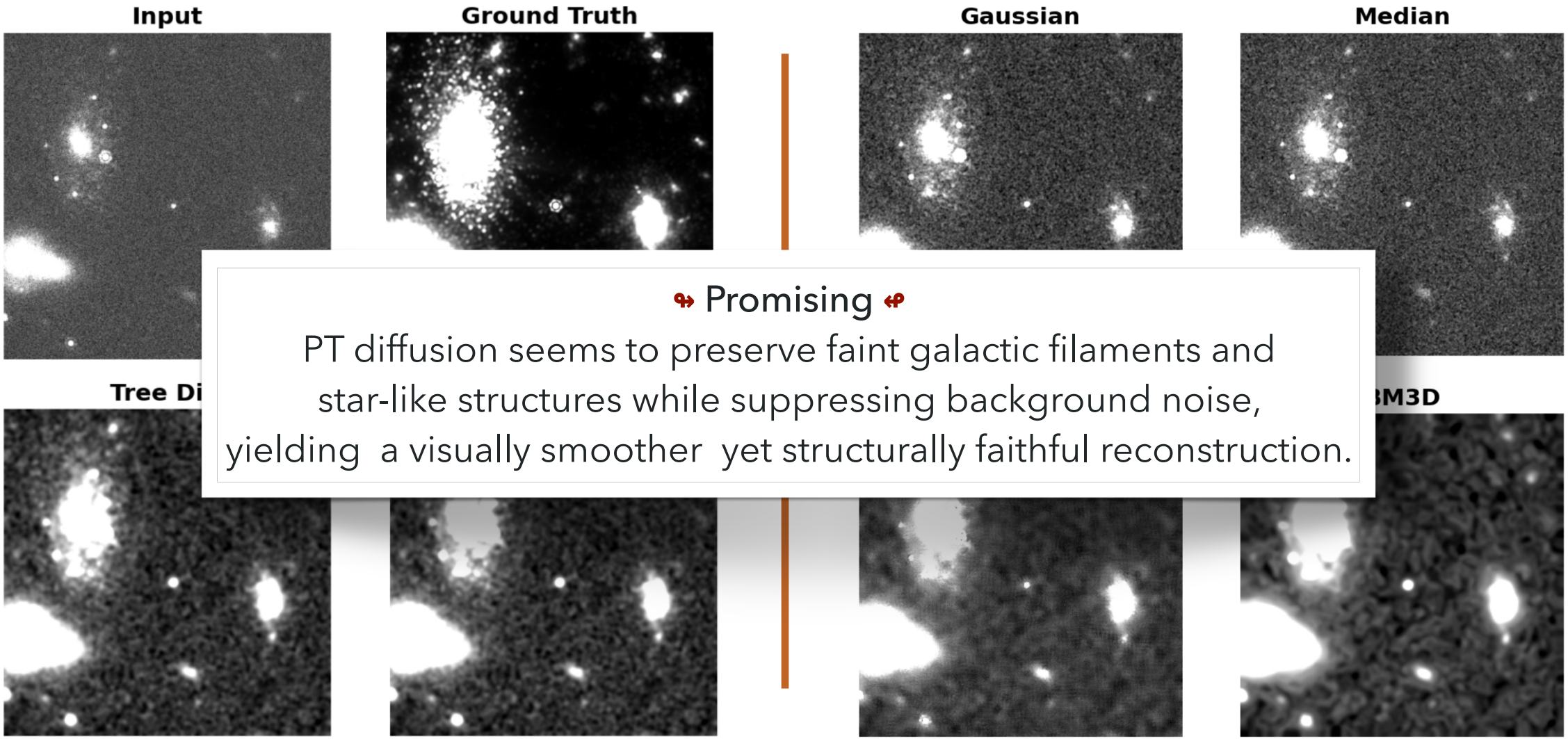


HIRDiff




Input
Ground Truth
Gaussian
Median
Tree Diff
Tree Diff (Cut)
NLMeans
BM3D

Finanziato dall'Unione europea
NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA
INAF
ISTITUTO NAZIONALE DI ASTROFISICA



Promising

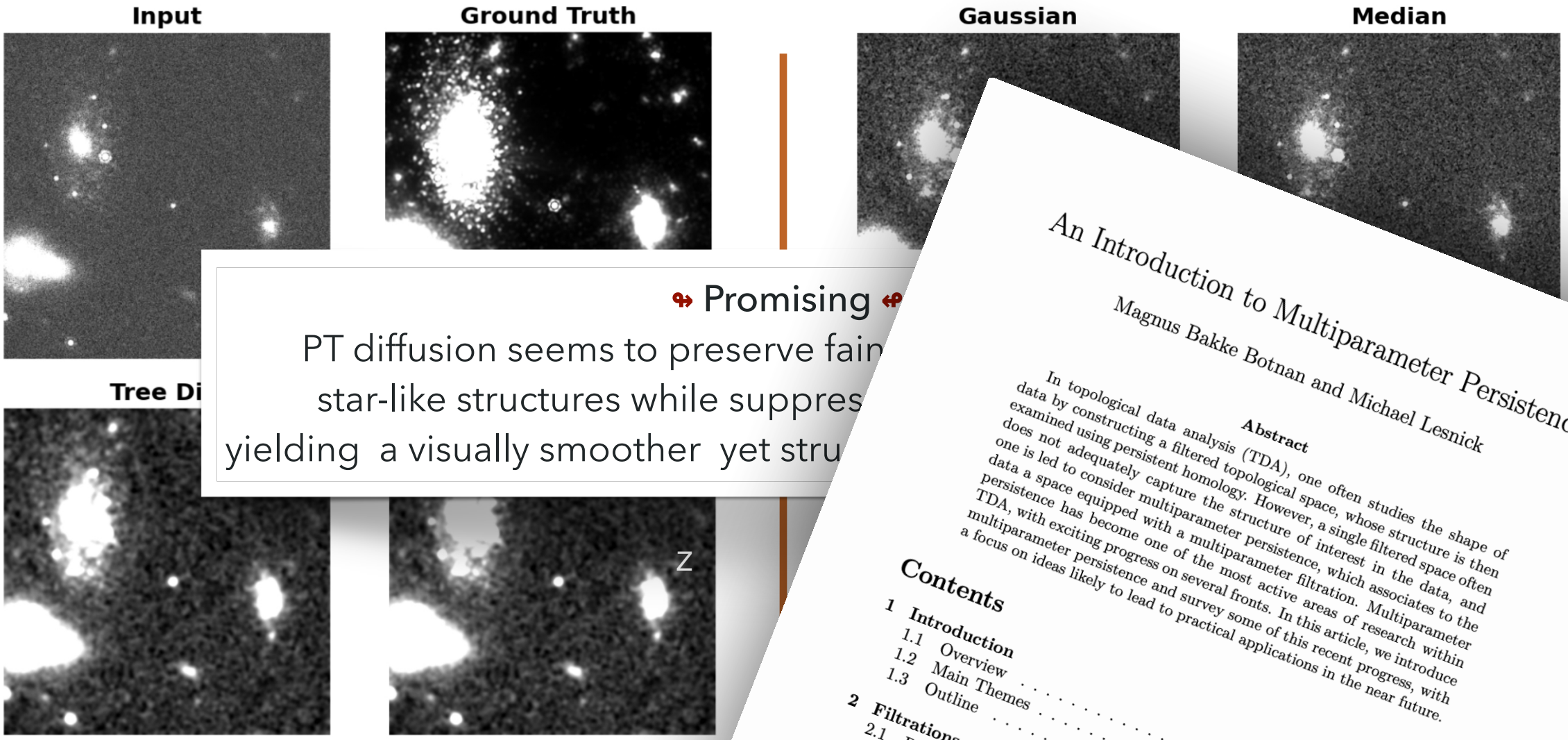
PT diffusion seems to preserve faint galactic filaments and star-like structures while suppressing background noise, yielding a visually smoother yet structurally faithful reconstruction.

Finanziato dall'Unione europea
NextGenerationEU

Ministero dell'Università e della Ricerca

Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA

INAF
ISTITUTO NAZIONALE DI ASTROFISICA



Promising

PT diffusion seems to preserve fain
star-like structures while suppres
yielding a visually smoother yet stru

An Introduction to Multiparameter Persistence

Magnus Bakke Botnan and Michael Lesnick

Abstract

In topological data analysis (TDA), one often studies the shape of data by constructing a filtered topological space, whose structure is then examined using persistent homology. However, a single filtered space often does not adequately capture the structure of interest in the data, and one is led to consider multiparameter persistence, which associates to the data a space equipped with a multiparameter filtration. Multiparameter persistence has become one of the most active areas of research within TDA, with exciting progress on several fronts. In this article, we introduce multiparameter persistence and survey some of this recent progress, with a focus on ideas likely to lead to practical applications in the near future.

Contents

1 Introduction
1.1 Overview
1.2 Main Themes
1.3 Outline
2 Filtrations and Persistence Modules
2.1 Posets
2.2 Filtrations
2.3 Persistence Mo
2.4 Persister
3

Finanziato dall'Unione europea
NextGenerationEU

Ministero dell'Università e della Ricerca

Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA

INAF
ISTITUTO NAZIONALE DI ASTROFISICA

...clearly, I'm no Steve Jobs either, but still...



PNRR
Missione 4 • Componente 2
Investimento 3.1

STILES – IR0000034
CUP C33C22000640006

Finanziato dall'Unione europea
NextGenerationEU

Ministero dell'Università e della Ricerca

Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA

INAF
ISTITUTO NAZIONALE DI ASTROFISICA

...clearly, I'm no Steve Jobs either, but still...



Finanziato dall'Unione europea
NextGenerationEU

Ministero dell'Università e della Ricerca

Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA

INAF
ISTITUTO NAZIONALE DI ASTROFISICA



. . . me staring at an astronomical image analysis pipeline . . .

PNRR
Missione 4 • Componente 2
Investimento 3.1

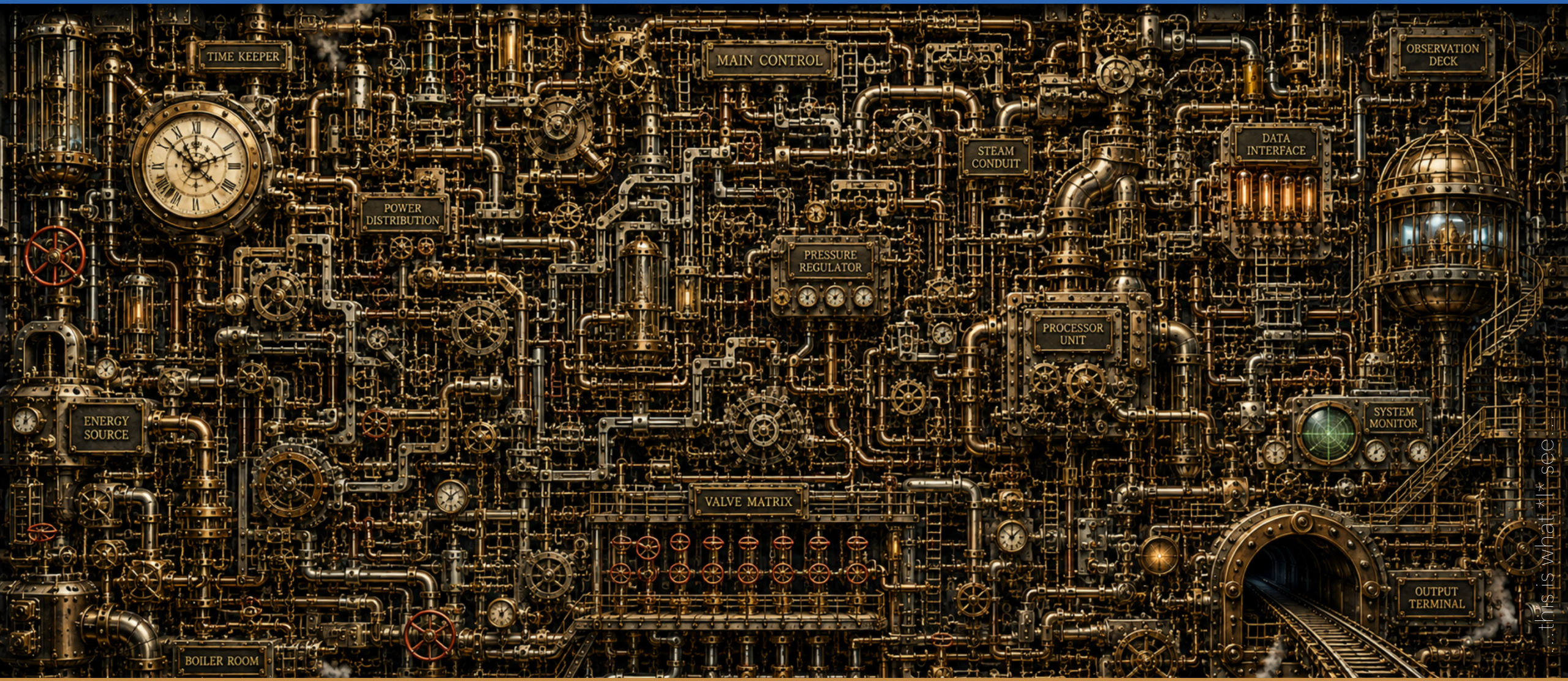
STILES – IR0000034
CUP C33C22000640006

Finanziato dall'Unione europea
NextGenerationEU

Ministero dell'Università e della Ricerca

Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA

INAF
ISTITUTO NAZIONALE DI ASTROFISICA



PNRR
Missione 4 • Componente 2
Investimento 3.1

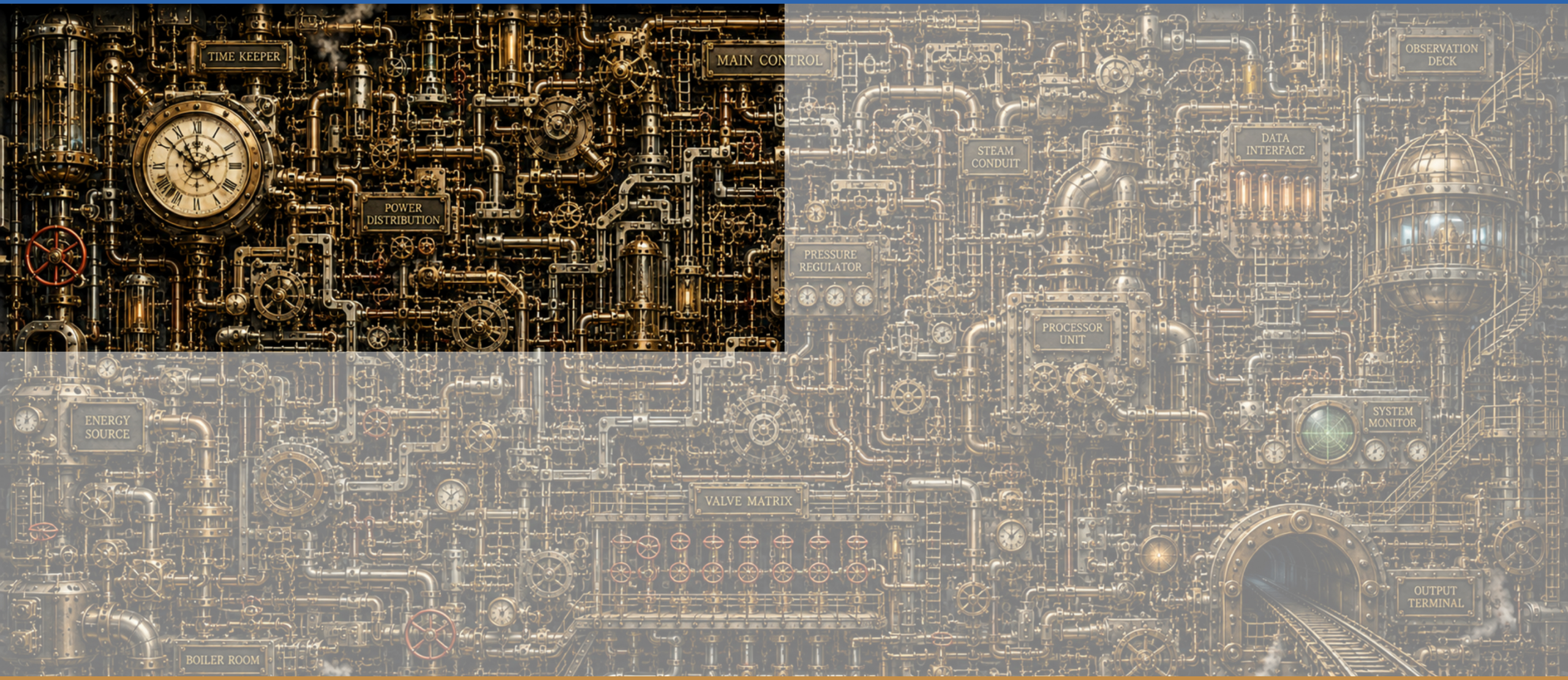
STILES – IR0000034
CUP C33C22000640006

Finanziato dall'Unione europea
NextGenerationEU

Ministero dell'Università e della Ricerca

Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA

INAF
ISTITUTO NAZIONALE DI ASTROFISICA

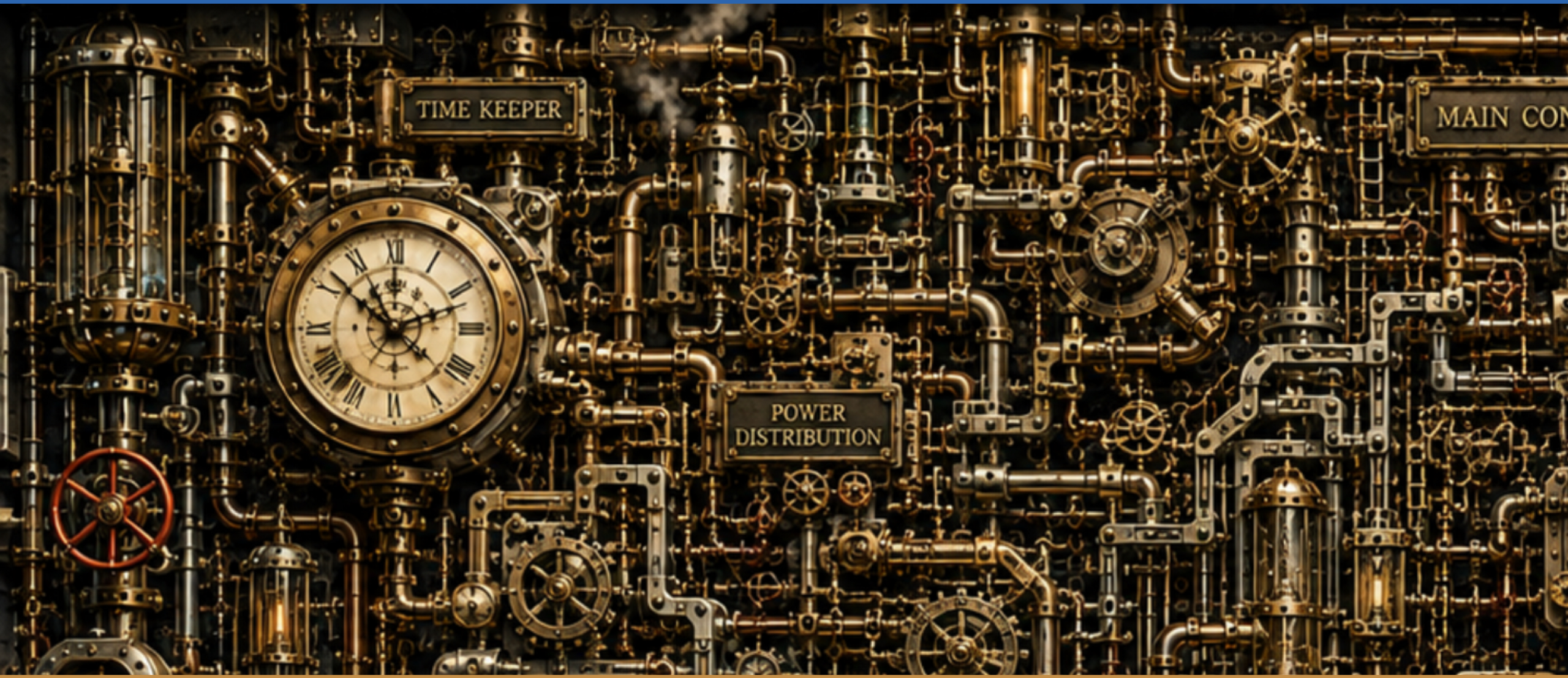


Finanziato dall'Unione europea
NextGenerationEU

Ministero dell'Università e della Ricerca

Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA

INAF
ISTITUTO NAZIONALE DI ASTROFISICA

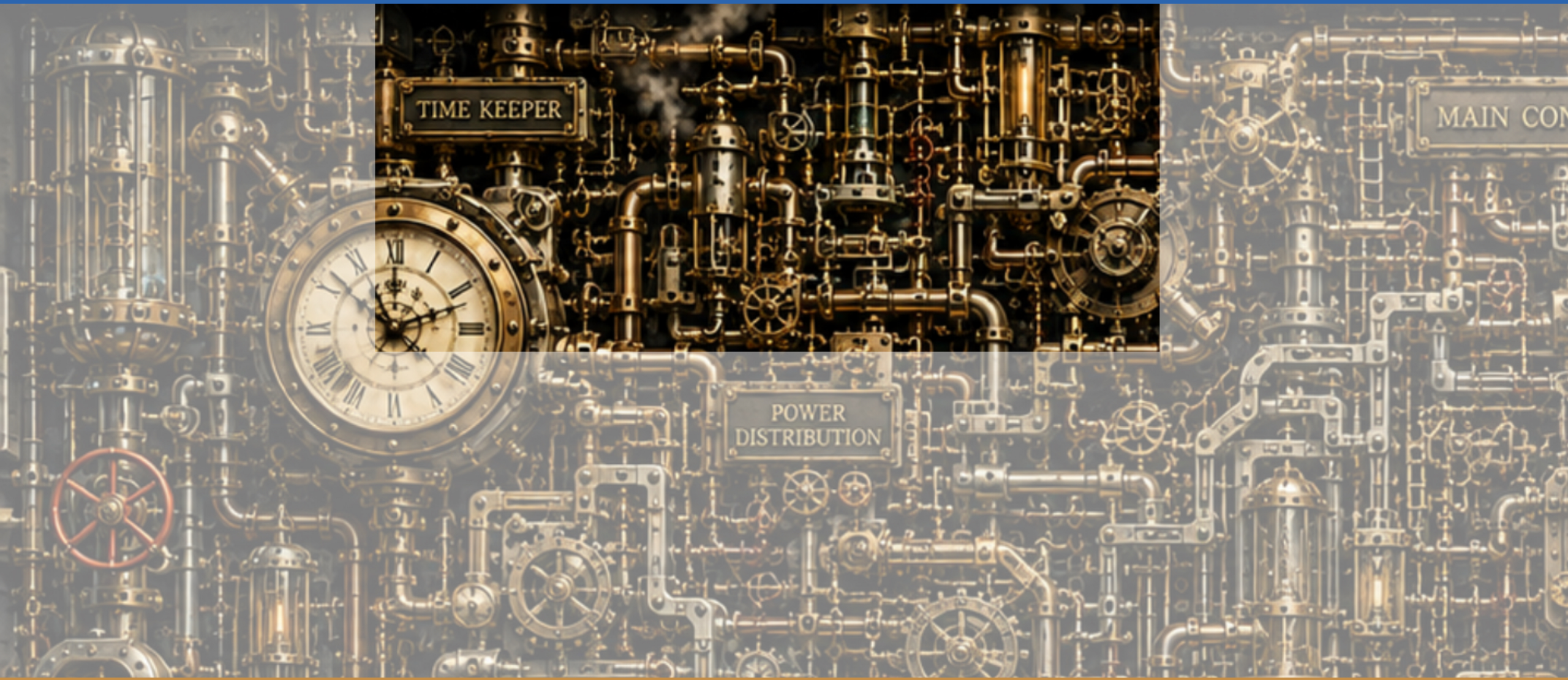


Finanziato dall'Unione europea
NextGenerationEU

Ministero dell'Università e della Ricerca

Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA

INAF
ISTITUTO NAZIONALE DI ASTROFISICA



Finanziato dall'Unione europea
NextGenerationEU

Ministero dell'Università e della Ricerca

Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA

INAF
ISTITUTO NAZIONALE DI ASTROFISICA



Finanziato dall'Unione europea
NextGenerationEU

Ministero dell'Università e della Ricerca

Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA

INAF
ISTITUTO NAZIONALE DI ASTROFISICA

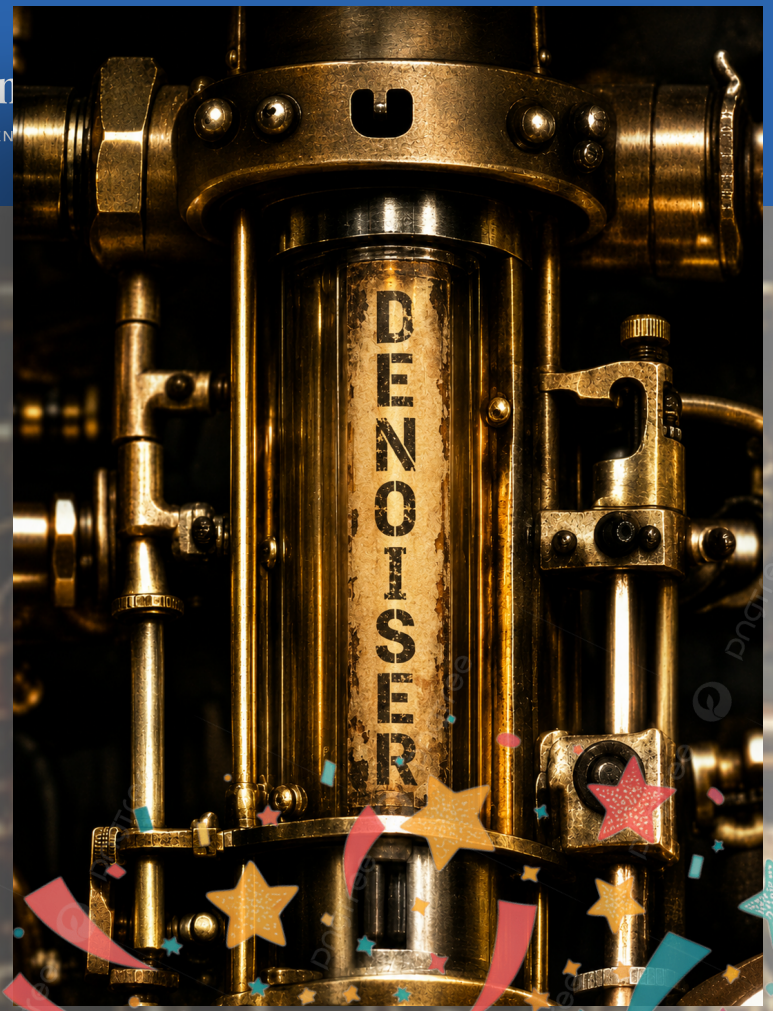


Finanziato dall'Unione europea
NextGenerationEU

Ministero dell'Università e della Ricerca

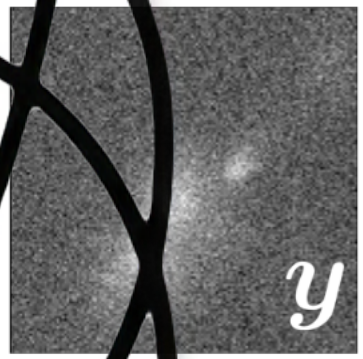
Italiadom
PIANO NAZIONALE DI RIPRESA E RESILIEN





Finanziato dall'Unione europea
NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA
INAF

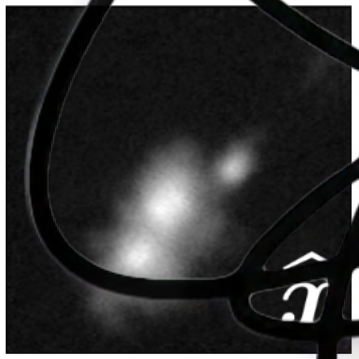
Noisy | Input signal



$$D(y, \sigma)$$

Image Denoiser

Denoised | Output signal



THE REAL BLACK-BOX
(…to me 😅)

PNRR
Missione 4 • Componente 2
Investimento 3.1
STILES – IR0000034
CUP C33C22000640006

Finanziato dall'Unione europea
NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA
INAF

My Medical Condition

2U2P

Beware! Endemic in the statistical community

PNRR
Missione 4 • Componente 2
Investimento 3.1
STILES – IR0000034
CUP C33C22000640006

Finanziato dall'Unione europea
NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA
INAF

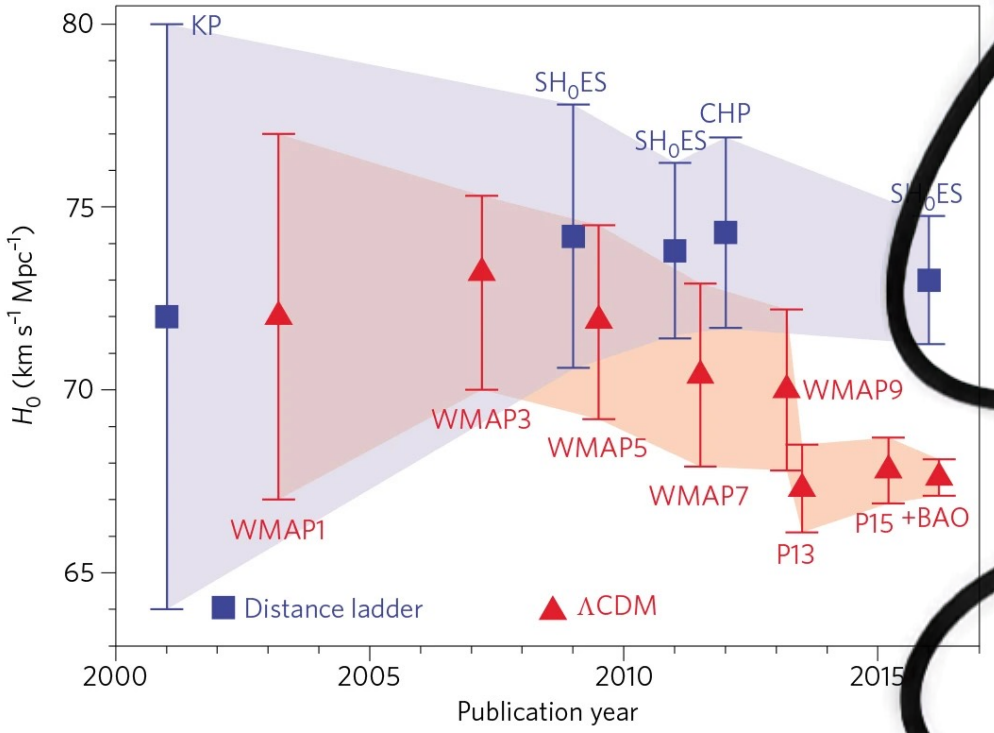
My Medical Condition

Uncertainty UnPropagation Phobia

Beware! Denoising by Sampling might help…or maybe not!

PNRR
Missione 4 • Componente 2
Investimento 3.1
STILES – IR0000034
CUP C33C22000640006

...and you **do** care about uncertainty!



Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
INAF ISTITUTO NAZIONALE DI ASTROFISICA

Idea

Conformally Robust Multistage Prediction

A Gentle Introduction to Conformal Prediction and Distribution-Free Uncertainty Quantification

Anastasios N. Angelopoulos and Stephen Bates

December 8, 2022

Abstract

ine learning models are now routinely used in high-risk settings, like uncertainty quantification to avoid consequential model failu inference) is a user-friendly paradigm for creating stati redictions of such models. Critically, the sets a n-asymptotic guarantees even without conformal prediction with any p anteed to contain the gr nd, easy-to-use, a ural langua

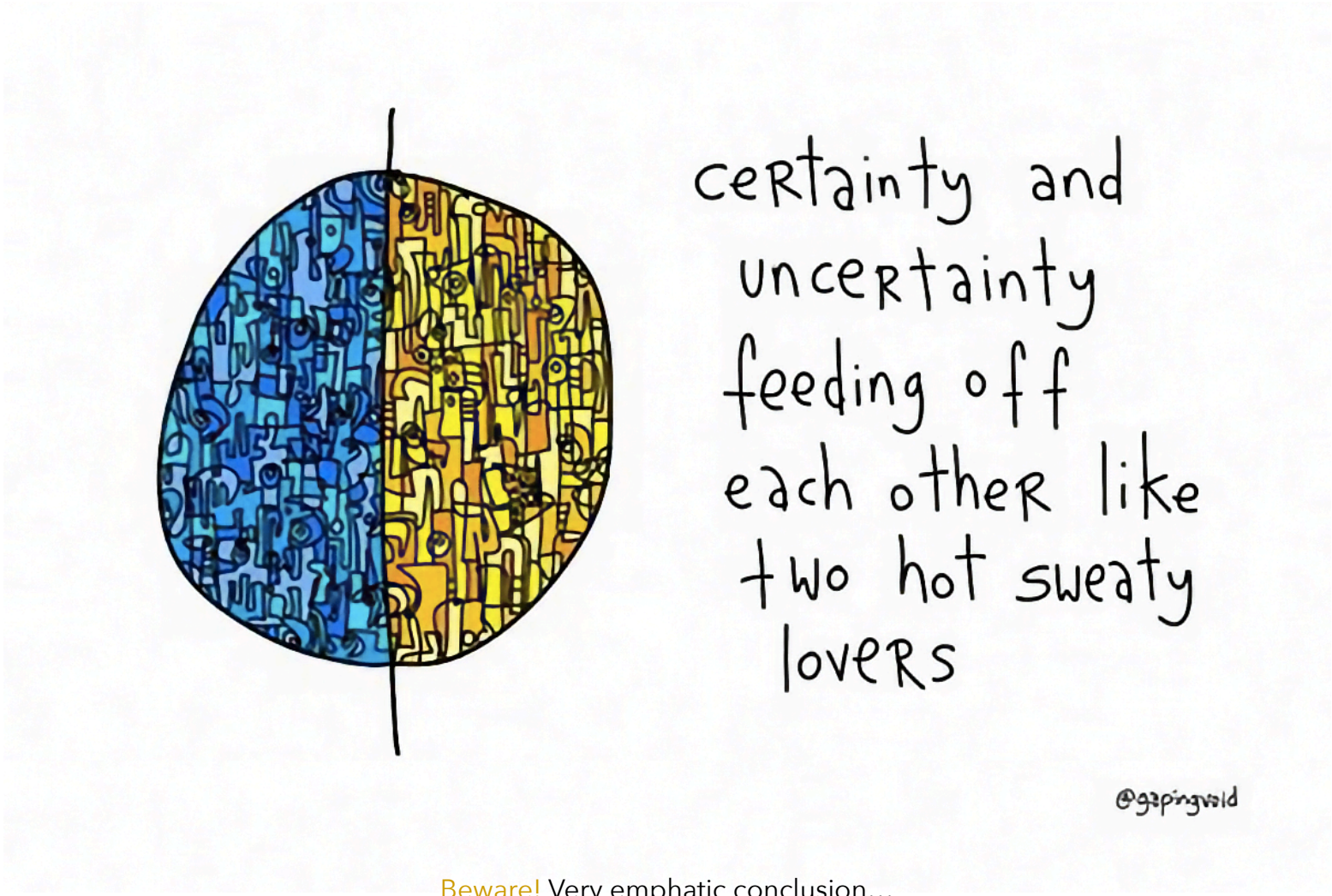
2008) 371-421

A Tutorial on Conformal Prediction

nd Information Systems

Submitted 8/07;

GSF





Beware! Very emphatic conclusion…

Finanziato dall'Unione europea
NextGenerationEU

Ministero dell'Università e della Ricerca

Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA

INAF
ISTITUTO NAZIONALE DI ASTROFISICA



https://teralab.ai/

Welcome to TeraLab

HPC for Statistics and AI

PNRR
Missione 4 • Componente 2
Investimento 3.1

STILES – IR0000034
CUP C33C22000640006

Finanziato dall'Unione europea
NextGenerationEU

Ministero dell'Università e della Ricerca

Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA

INAF
ISTITUTO NAZIONALE DI ASTROFISICA









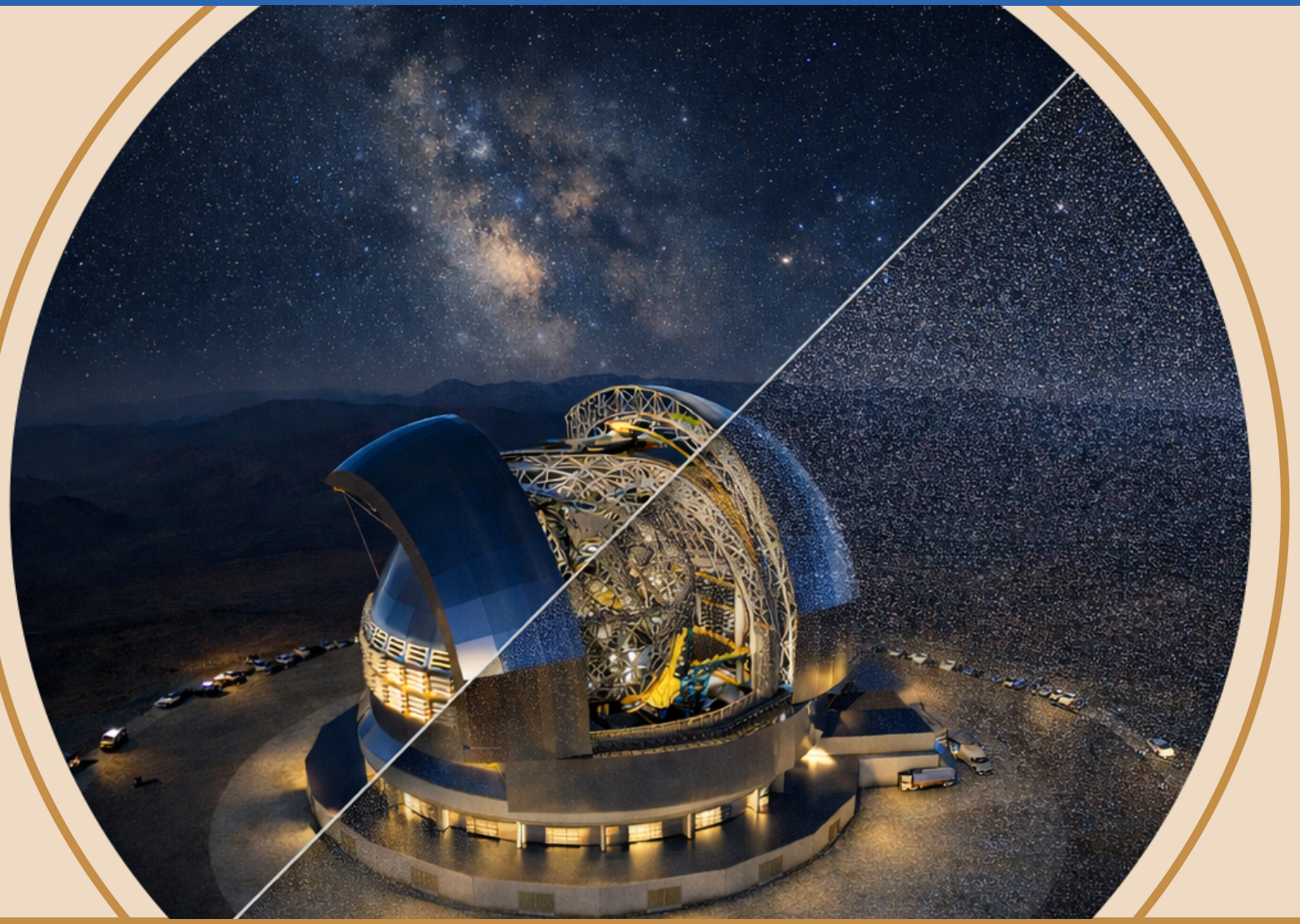








Thanks!



*See (some of) you in a future project…maybe ^.^)