Generalizability (or not?) of patch-based image models

Jeffrey A. Fessler & Jason Hu

EECS Department, BME Department, Dept. of Radiology
University of Michigan

http://web.eecs.umich.edu/~fessler

BASP Frontiers conference
2023-02-07

Acknowledgments:
Xiaojian Xu, Mike McCann (LANL)

Outline

Introduction
- Generative models

Patch-based score modeling

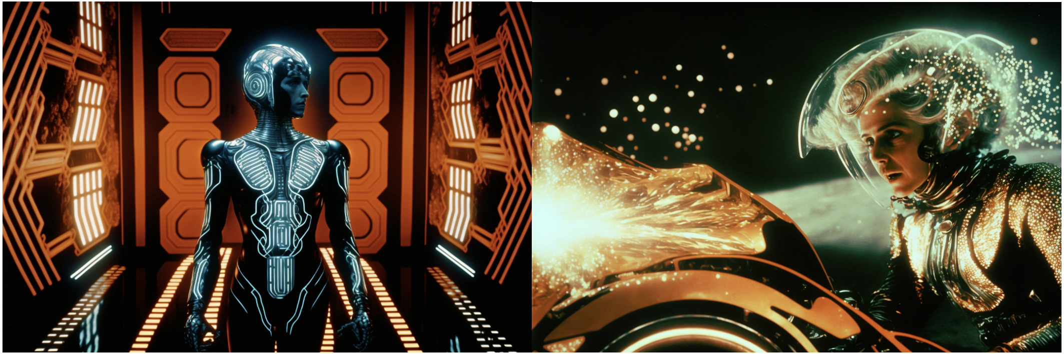
Current results

Summary

Bibliography

Extra: toy exploration

Generative models are hot in graphics



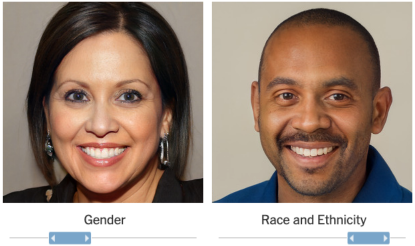
Computer ("AI") generated stills from hypothetical movie: Chilean director Alejandro Jodorowsky's 1976 version of "Tron" using midjourney.com as reported in 2023-01-13 NY Times article "This film does not exist" by director Frank Pavich.

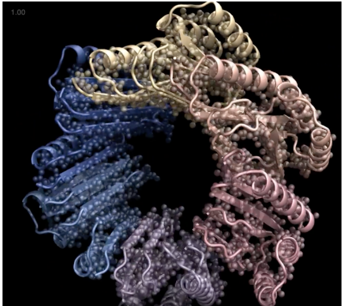
Generative models are hot in the news

- 2020-11-21 NY Times "Designed to Deceive: Do These People Look Real to You?"
  Article about generated (aka fake) faces.

- 2022-10-21 NY Times "A Coming-Out Party for Generative A.I., Silicon Valley's New Craze"
  (about "Stable Diffusion" image generator)
  `https://nyti.ms/3SjsN0k`

- 2023-01-09 NY Times "A.I. Turns Its Artistry to Creating New Human Proteins"
  `https://nyti.ms/3IzY66m`





Generative models are hot in imaging / inverse problems

Zhao, Ye, Bresler: Jan. 2023 IEEE SpMag survey paper [1]

- Generative adversarial network (GAN) models
- Variation auto-encoder (VAE) models [2]
- Normalizing flows [3]
- Score-based diffusion models
  - Ramzi et al., NeurIPS 2020 [4]
  - Yang Song et al., NeurIPS 2021, ICLR 2022 [5, 6]
  - Jalal et al., NeurIPS 2021 [7]
  - Chung & Ye, MIA, Aug. 2022 [8]
  - Luo et al., 2022 arXiv 2202.01479 [9]
  - ...
- Kazerouni et al. [10] have github catalog, including 5 survey papers
- ... (hopelessly incomplete lists)

Benefits of generative models

- Can learn prior $\mathrm{p}(\boldsymbol{x})$ independent of system, *e.g.*, MRI k-space sampling patterns (Though may depend on pixel size and contrast.)
- Unsupervised learning [11]
(Though reasonably high-quality training data may be needed.)
- Given image prior $\mathrm{p}(\boldsymbol{x})$, can use Bayes rule to sample from the posterior $\mathrm{p}(\boldsymbol{x}|\boldsymbol{y})$ for uncertainty quantification (recent survey: [12])
- Sampling needs just its score function $\nabla \log \mathrm{p}(\boldsymbol{x}; \boldsymbol{\theta})$,
using Langevin dynamics, aka stochastic gradient ascent of log-likelihood:

$$\boldsymbol{x}_t = \boldsymbol{x}_{t-1} + \alpha_t \nabla \log \mathrm{p}(\boldsymbol{x}_{t-1}) + \beta_t \mathcal{N}(\mathbf{0}, \boldsymbol{I}), \quad t = 1, \ldots, T.$$

  - Draws samples from $\mathrm{p}(\boldsymbol{x})$ for suitable choices of $\{\alpha_t\}$, $\{\beta_t\}$, and (large) T [13].
  - (See [14] for acceleration for inverse problems using data consistency.)

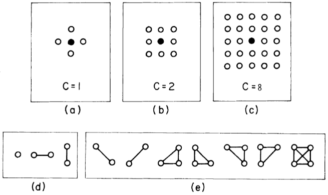
NY Times article
about fake faces

See it?

Long history of generative models and inverse problems

Markov random field models

(*e.g.*) Geman & Geman 1984 [15]



Mostly for inference?

GEMAN AND GEMAN: STOCHASTIC RELAXATION, GIBBS DISTRIBUTIONS, AND BAYESIAN RESTORATION 737

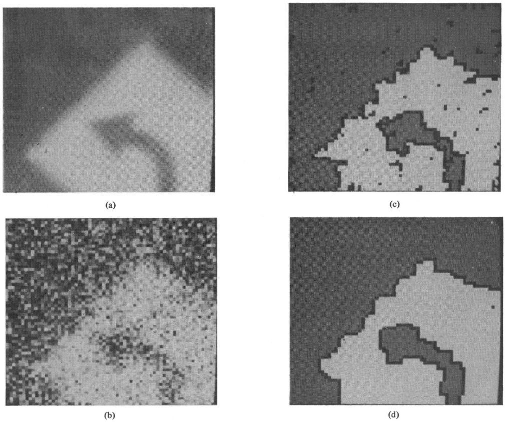


Fig. 7. (a) Blurred image (roadside scene). (b) Degraded image: Additive noise. (c) Restoration including line process; 100 iterations. (d) Restoration including line process; 1000 iterations.

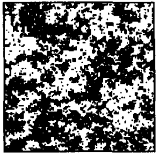
Long history of generative models and inverse problems

MRF as generators? [16] T-PAMI 1994

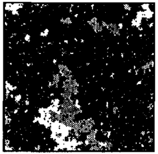
An Empirical Study of the Simulation of Various Models Used for Images

A. J. Gray, J. W. Kay, and D. M. Titterington

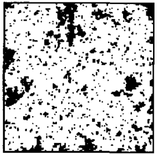
***Abstract*— Markov random fields are typically used as priors in Bayesian image restoration methods to represent spatial information in the image. Commonly used Markov random fields are not in fact capable of representing the moderate-to-large scale clustering present in naturally occurring images and can also be time consuming to simulate,**



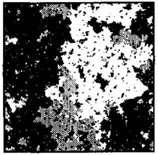
(b)



(f)



(c)

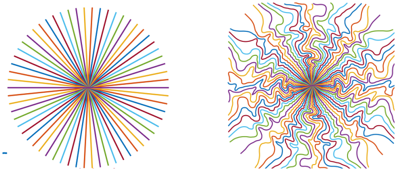


(g)

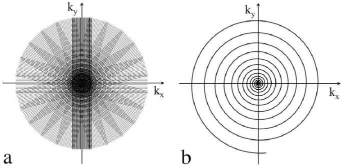
Whole images vs patches?

Jan. 2023 survey paper on generative models [1] does not mention “patch” once!?

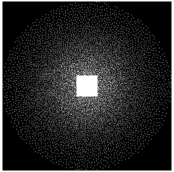
MRI k-space sampling:



[17]



[18]



[19]

Patch-based models have long history in inverse problems, *e.g.*,

- patch GAN [20–22]
- patch dictionary models [23, 24]
- non-local means, BM3D ...

Motivating questions

- Could patch-based generative models provide better robustness to distribution shifts, perhaps at the cost of reduced in-distribution performance?
- Especially in applications with very limited training data?
  e.g., dynamic MRI
- Can we use the "latest" generative models, namely score-based models, for patches?

Patch-based score modeling

- Start with MRF formulation, aka *product of experts* model [25]:

$$\mathrm{p}(\boldsymbol{x};\boldsymbol{\theta}) = \frac{1}{Z(\boldsymbol{\theta})}\,\mathrm{e}^{-\sum_c V_c(\boldsymbol{x};\boldsymbol{\theta})} = \frac{1}{Z(\boldsymbol{\theta})}\prod_c \mathrm{e}^{-V_c(\boldsymbol{x};\boldsymbol{\theta})}.$$

  - $\boldsymbol{\theta}$: parameter vector that describes the prior
  - V_c : *clique potential* for the cth image *patch*
  - $Z(\boldsymbol{\theta})$: intractable partition function
- Assume statistical spatial stationarity (image shift invariance):

$$V_c(\boldsymbol{x};\boldsymbol{\theta}) = V(\boldsymbol{G}_c\boldsymbol{x};\boldsymbol{\theta}),$$

  - $\boldsymbol{G}_c$: wide binary matrix that grabs pixels of the cth patch from image $\boldsymbol{x}$
  - $V(\boldsymbol{z};\boldsymbol{\theta})$: common parent clique function

Patch-based score modeling

- Resulting log-prior:

$$\log \mathrm{p}(\boldsymbol{x}; \boldsymbol{\theta}) = -\log Z(\boldsymbol{\theta}) - \sum_c V(\boldsymbol{G}_c \boldsymbol{x}; \boldsymbol{\theta})$$

- Corresponding overall *image score function* arises from *patch score function*:

$$\boldsymbol{s}(\boldsymbol{x}; \boldsymbol{\theta}) \triangleq \nabla_{\boldsymbol{x}} \log \mathrm{p}(\boldsymbol{x}; \boldsymbol{\theta}) = -\sum_c \boldsymbol{G}_c' \boldsymbol{s}_V(\boldsymbol{G}_c \boldsymbol{x}; \boldsymbol{\theta}), \qquad \boldsymbol{s}_V(\boldsymbol{v}; \boldsymbol{\theta}) \triangleq \nabla_{\boldsymbol{v}} V(\boldsymbol{v}; \boldsymbol{\theta}).$$

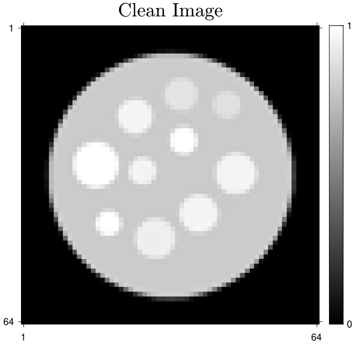
- All we must learn is the patch score function $\boldsymbol{s}_V(\boldsymbol{v}; \boldsymbol{\theta}) : \mathbb{R}^n \mapsto \mathbb{R}^n$, *e.g.*, a MLP.
- For training image patches $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_T\}$, apply *denoising score matching* (DSM) of Vincent, 2011 [26], typically for a range of noise variances σ^2 [13]:

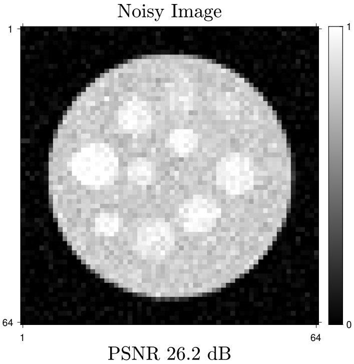
$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \frac{1}{T} \sum_{t=1}^{T} \mathrm{E}_{\sigma \sim p(\sigma)} \left[\sigma^2 \, \mathrm{E}_{\boldsymbol{z} \sim \mathcal{N}(0, \sigma^2 \boldsymbol{I}_n)} \left[\frac{1}{2} \left\| \boldsymbol{s}_V(\boldsymbol{v}_t + \boldsymbol{z}; \boldsymbol{\theta}, \sigma) + \frac{\boldsymbol{z}}{\sigma^2} \right\|_2^2 \right] \right].$$

- Final patch score model is $\boldsymbol{s}_V(\boldsymbol{v}; \hat{\boldsymbol{\theta}}, \sigma_{\min})$.

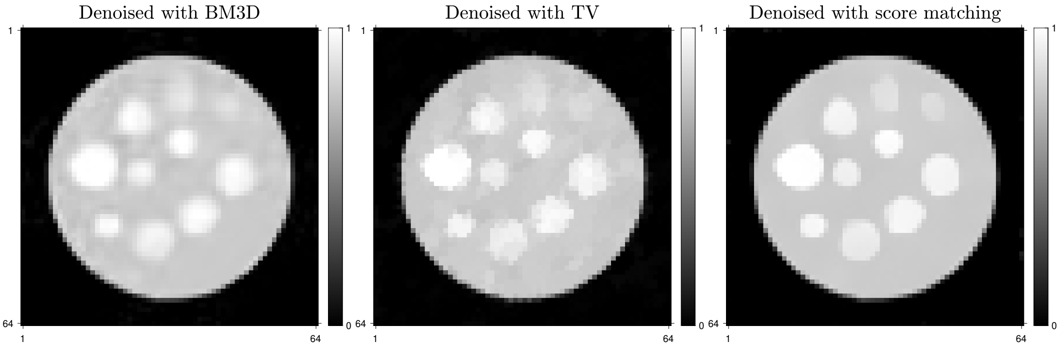
Simple exploration with *anecdotal* results

- 3×3 patches
- MLP patch score model (9, 40, 80, 160, 320, 320, 160, 80, 40, 9)
  first 5 with leaky ReLU, last 3 with tanh
- 1000 similar training examples





Denoising results



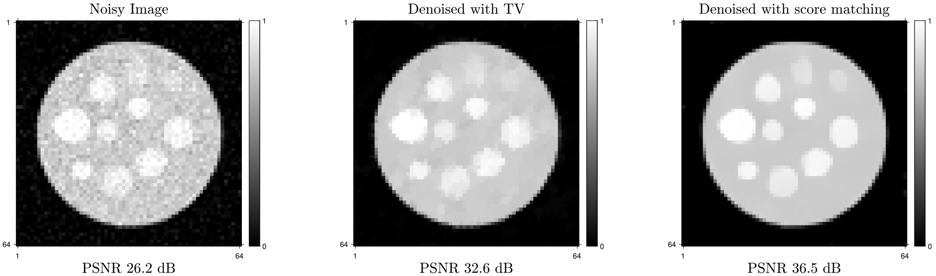
PSNR 34.0 dB — PSNR 32.6 dB — PSNR 36.5 dB

- TV regularization parameter optimized by oracle for best PSNR.
- MAP estimate by greedy gradient ascent of log posterior: (no β!)

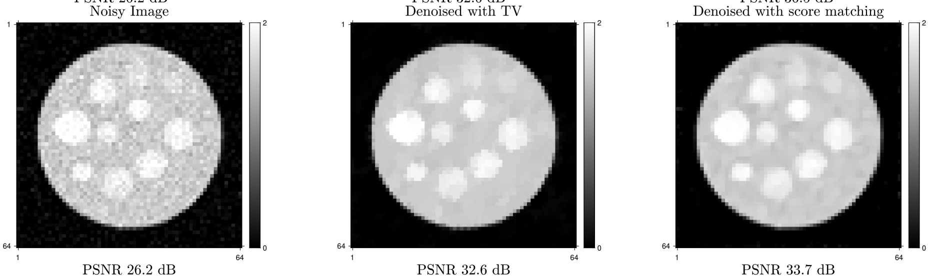
$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k \nabla_{\boldsymbol{x}} \log \mathrm{p}\left(\boldsymbol{x}_k | \boldsymbol{y}; \hat{\boldsymbol{\theta}}\right) = \boldsymbol{x}_k + \alpha_k \left(\nabla_{\boldsymbol{x}} \log \mathrm{p}(\boldsymbol{y} | \boldsymbol{x}_k) - \sum_c \boldsymbol{G}_c' \boldsymbol{s}_V(\boldsymbol{G}_c \boldsymbol{x}_k; \hat{\boldsymbol{\theta}})\right).$$

Generalizability to distribution shift? (pitfalls...)

Old:

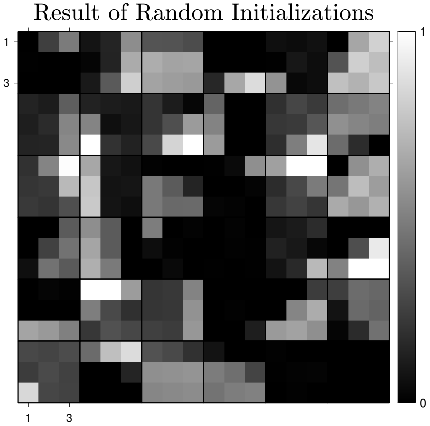


New:



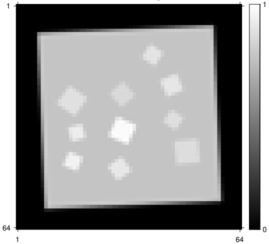
What changed?

MAP from random noise

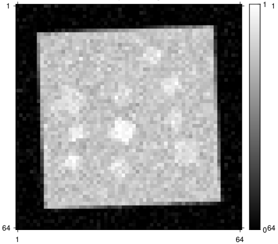


Distribution shift: rectangle test image

Clean Image

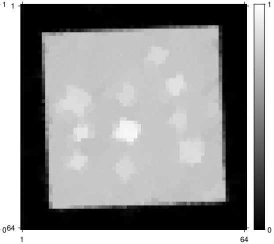


Noisy Image



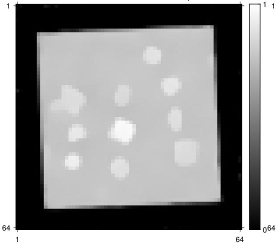
PSNR 26.0 dB

Denoised with TV



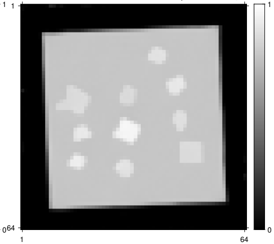
PSNR 33.2 dB

Score denoised, trained w/circles



PSNR 35.6 dB

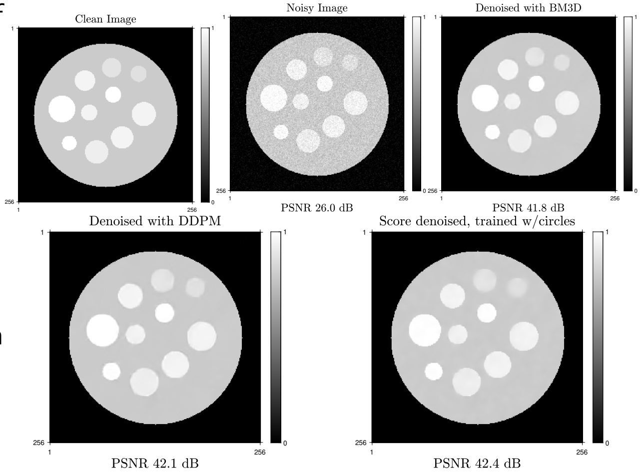
Score Denoised, Trained w/Rectangles



PSNR 37.7 dB

Whole-image vs patch models

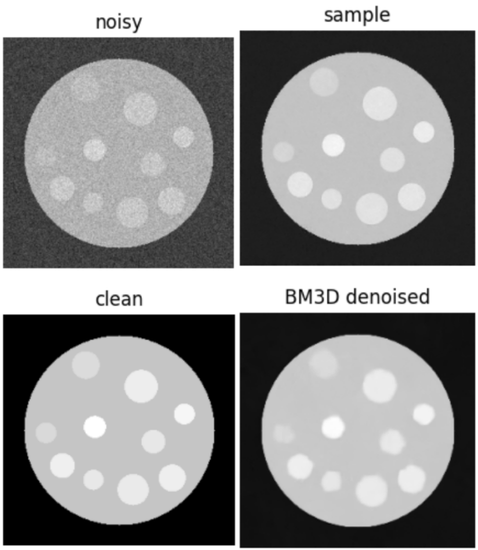
- Whole-image diffusion model of Hu et al. (SPIE, 2022) [27]
- `https://github.com/DeweiHu/OCT_DDPM`
- Based on Ho et al. (NeurIPS, 2020) [28] denoising diffusion prob. model (DDPM)
- Trained with 1000 disk images.
- Tested with noisy disk phantom
- One sample from posterior



Whole-image models and generalizability?

- Diffusion model of Hu et al. (SPIE, 2022) [27] trained with 3600 flower images.

- Tested with noisy disk phantom (PSNR 20.3 dB)
- One sample from posterior `https://github.com/DeweiHu/OCT_DDPM`



37.0 dB

35.3 dB

Summary / future directions

- Learning patch score models is feasible with denoising score matching
- Amplitude scale invariance is not inherent to score-based models
  Easily (?) fixed by patch normalization, but what other more subtle pitfalls exist?
- Integrate invariances: amplitude scale / rotation / flip / DC offset
- Compare with whole-image models:
  - "pure" CNN score models with small receptive fields
  - multi-scale score models [29, 30]
  - ...
- Explore trade-offs between generalizability and in-distribution performance
- Is the "optimal" patch size the whole image? (Even for 3D+T?)

Tutorial Julia code: `https://github.com/JeffFessler/ScoreMatching.jl`

Resources

Talk and code available online at
`http://web.eecs.umich.edu/~fessler`

Bibliography I

[1] Z. Zhao, J. C. Ye, and Y. Bresler. "Generative models for inverse imaging problems: from mathematical foundations to physics-driven applications." In: *IEEE Sig. Proc. Mag.* 40.1 (2023), 148–63.

[2] E. D. Zhong, T. Bepler, B. Berger, and J. H. Davis. "CryoDRGN: reconstruction of heterogeneous cryo-EM structures using neural networks." In: *Nature Meth.* 18.2 (2021), 176–85.

[3] D. Rezende and S. Mohamed. "Variational inference with normalizing flows." In: *Proc. Intl. Conf. Mach. Learn.* 2015, 1530–8.

[4] Z. Ramzi, B. Remy, F. Lanusse, J-L. Starck, and P. Ciuciu. "Denoising score-matching for uncertainty quantification in inverse problems." In: *NeurIPS 2020 Workshop on Deep Learning and Inverse Problems*. 2020.

[5] Y. Song, L. Shen, L. Xing, and S. Ermon. "Solving inverse problems in medical imaging with score-based generative models." In: *NeurIPS Deep Inv. Work.* 2021.

[6] Y. Song, L. Shen, L. Xing, and S. Ermon. "Solving inverse problems in medical imaging with score-based generative models." In: *Proc. Intl. Conf. on Learning Representations*. 2022.

[7] A. Jalal, M. Arvinte, G. Daras, E. Price, A. Dimakis, and J. Tamir. "Robust compressed sensing MR imaging with deep generative priors." In: *NeurIPS Workshop Deep Inverse*. 2021.

[8] H. Chung and J. C. Ye. "Score-based diffusion models for accelerated MRI." In: *Med. Im. Anal.* 80 (Aug. 2022), p. 102479.

[9] G. Luo, M. Heide, and M. Uecker. *MRI reconstruction via data driven Markov chain with joint uncertainty estimation*. 2022.

[10] A. Kazerouni, E. K. Aghdam, M. Heidari, R. Azad, M. Fayyaz, I. Hacihaliloglu, and D. Merhof. *Diffusion models for medical image analysis: A comprehensive survey*. 2022.

[11] M. Akcakaya, B. Yaman, H. Chung, and J. C. Ye. "Unsupervised deep learning methods for biological image reconstruction and enhancement: an overview from a signal processing perspective." In: *IEEE Sig. Proc. Mag.* 39.2 (2022), 28–44.

[12] W. Dong, J. Wu, L. Li, G. Shi, and X. Li. "Bayesian deep learning for image reconstruction: from structured sparsity to uncertainty estimation." In: *IEEE Sig. Proc. Mag.* 40.1 (2023), 73–84.

Bibliography II

[13] Y. Song, J. Sohl-Dickstein, D. P. Kingma, A. Kumar, S. Ermon, and B. Poole. "Score-based generative modeling through stochastic differential equations." In: *Proc. Intl. Conf. on Learning Representations*. 2021.

[14] H. Chung, B. Sim, and J. C. Ye. "Come-closer-diffuse-faster: accelerating conditional diffusion models for inverse problems through stochastic contraction." In: *Proc. IEEE Conf. on Comp. Vision and Pattern Recognition*, 12403–12.

[15] S. Geman and D. Geman. "Stochastic relaxation, Gibbs distributions, and Bayesian restoration of images." In: *IEEE Trans. Patt. Anal. Mach. Int.* 6.6 (Nov. 1984), 721–41.

[16] A. J. Gray, J. W. Kay, and D. M. Titterington. "An empirical study of the simulation of various models used for images." In: *IEEE Trans. Patt. Anal. Mach. Int.* 16.5 (May 1994), 507–12.

[17] G. Wang, T. Luo, J-F. Nielsen, D. C. Noll, and J. A. Fessler. "B-spline parameterized joint optimization of reconstruction and k-space trajectories (BJORK) for accelerated 2D MRI." In: *IEEE Trans. Med. Imag.* 41.9 (Sept. 2022), 2318–30.

[18] W. Wu and K. L. Miller. "Image formation in diffusion MRI: A review of recent technical developments." In: *J. Mag. Res. Im.* 46.3 (Sept. 2017), 646–62.

[19] S. Bhadra, W. Zhou, and M. A. Anastasio. "Medical image reconstruction with image-adaptive priors learned by use of generative adversarial networks." In: *Proc. SPIE 11312 Medical Imaging: Phys. Med. Im.* 2020, p. 113120V.

[20] C. Li and M. Wand. "Precomputed real-time texture synthesis with Markovian generative adversarial networks." In: *Proc. European Comp. Vision Conf.* 2016, 702–16.

[21] P. Isola, J-Y. Zhu, T. Zhou, and A. A. Efros. "Image-to-image translation with conditional adversarial networks." In: *Proc. IEEE Conf. on Comp. Vision and Pattern Recognition*. 2017, 5967–76.

[22] A. Elnekave and Y. Weiss. "Generating natural images with direct patch distributions matching." In: *Proc. European Comp. Vision Conf.* Vol. 13677. 2022.

[23] M. Aharon, M. Elad, and A. Bruckstein. “K-SVD: an algorithm for designing overcomplete dictionaries for sparse representation.” In: *IEEE Trans. Sig. Proc.* 54.11 (Nov. 2006), 4311–22.

[24] S. Ravishankar and Y. Bresler. “MR image reconstruction from highly undersampled k-space data by dictionary learning.” In: *IEEE Trans. Med. Imag.* 30.5 (May 2011), 1028–41.

[25] G. E. Hinton. “Training products of experts by minimizing contrastive divergence.” In: *Neural Computation* 14.8 (Aug. 2002), 1771–800.

[26] P. Vincent. “A connection between score matching and denoising autoencoders.” In: *Neural Comput.* 23.7 (July 2011), 1661–74.

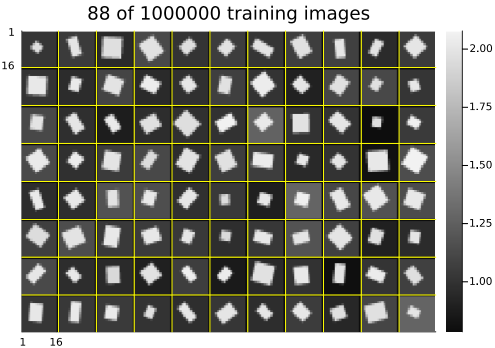
[27] D. Hu, Y. K. Tao, and I. Oguz. “Unsupervised denoising of retinal OCT with diffusion probabilistic model.” In: *spie-12032*. 2022, p. 1203206.

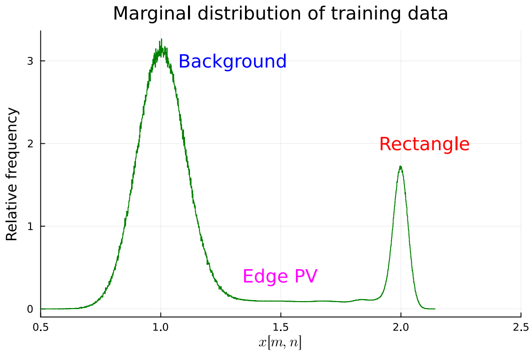
[28] J. Ho, A. Jain, and P. Abbeel. “Denoising diffusion probabilistic models.” In: *NeurIPS*. Vol. 33. 2020, 6840–51.

[29] F. Guth, S. Coste, V. D. Bortoli, and Stéphane Mallat. “Wavelet score-based generative modeling.” In: *NeurIPS*. 2022.

[30] Z. Kadkhodaie, F. Guth, Stephane Mallat, and E. P. Simoncelli. “Learning multi-scale local conditional probability models of images.” In: *Proc. Intl. Conf. on Learning Representations*. 2023.

A simple exploration

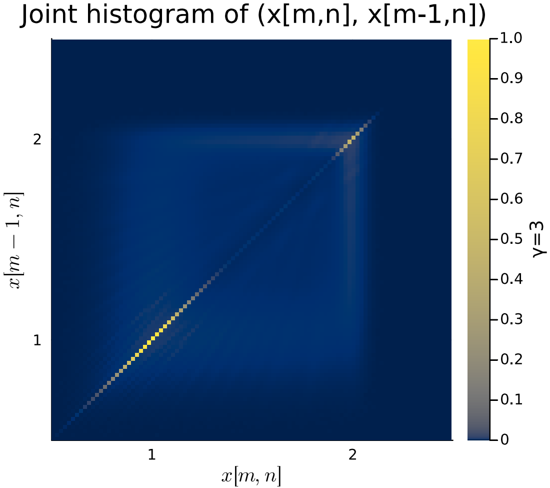




- Stochastic image model with random: center, width, orientation, background $\mathcal{N}(1, 0.1^2)$, rectangle foreground $\mathcal{N}(1, 0.03^2)$
- 10^6 training images of size 16×16 with partial volume effects.
- Data lies on 7-dimensional manifold.

Patch statistics: joint distribution

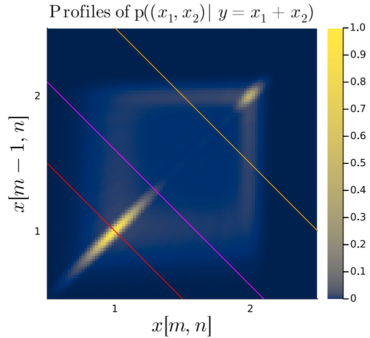
2×1 patches (cf TV)

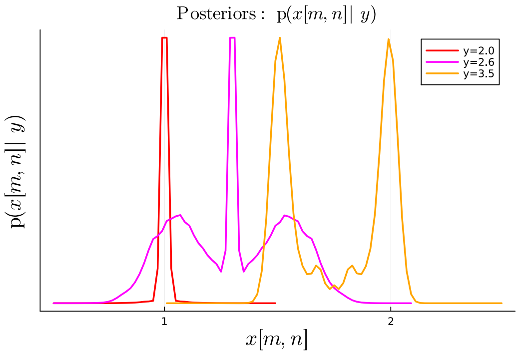


Patch statistics: posterior distributions

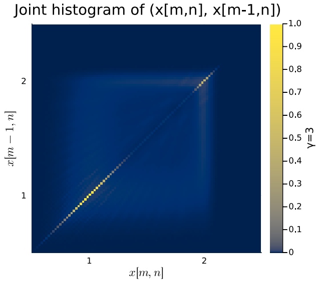
$$\mathrm{p}((x[m,n], x[m,n-1]) \mid y = x[m,n] + x[m,n-1])$$

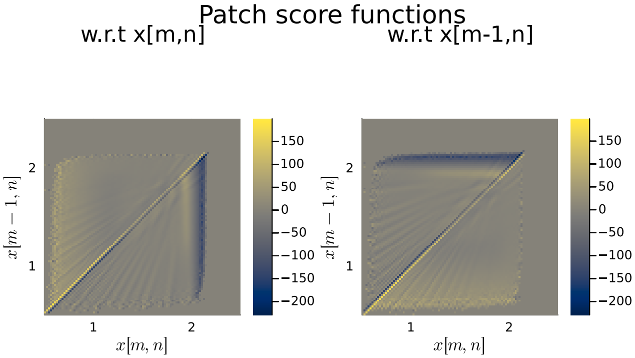
- MRI "center of k-space"
- MRI "2× acceleration





Patch statistics: score functions



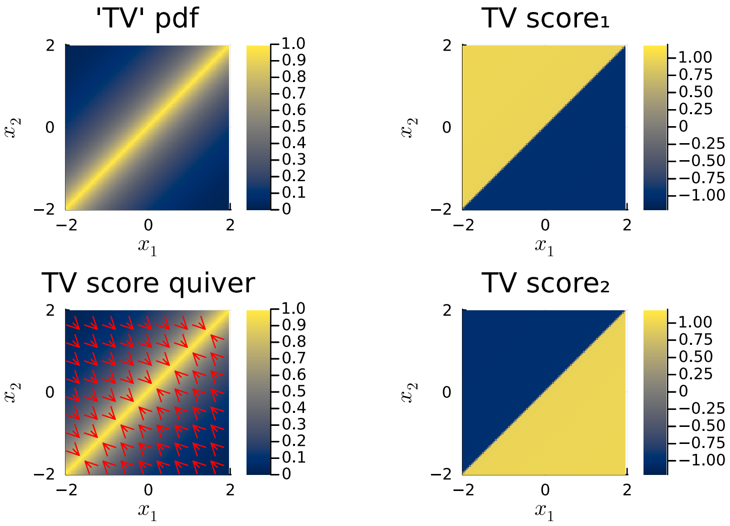


(Manifold data $\Longrightarrow$ score function $\boldsymbol{s}(\boldsymbol{x}) = \nabla_{\boldsymbol{x}} \log \mathrm{p}(\boldsymbol{x})$ is not well-defined.)

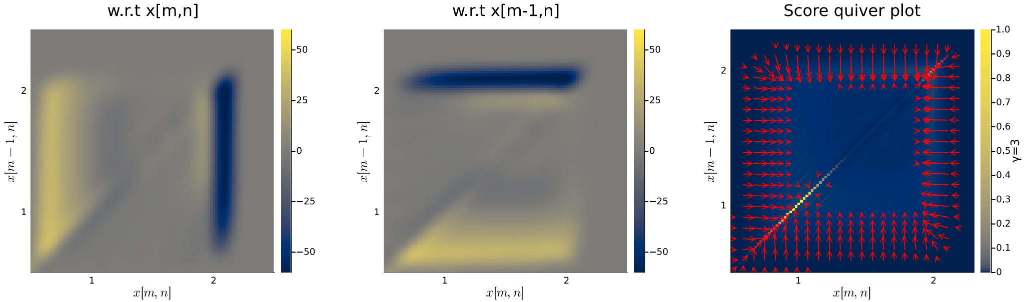
“TV” score function

Total variation (TV) prior for 2×1 patch:

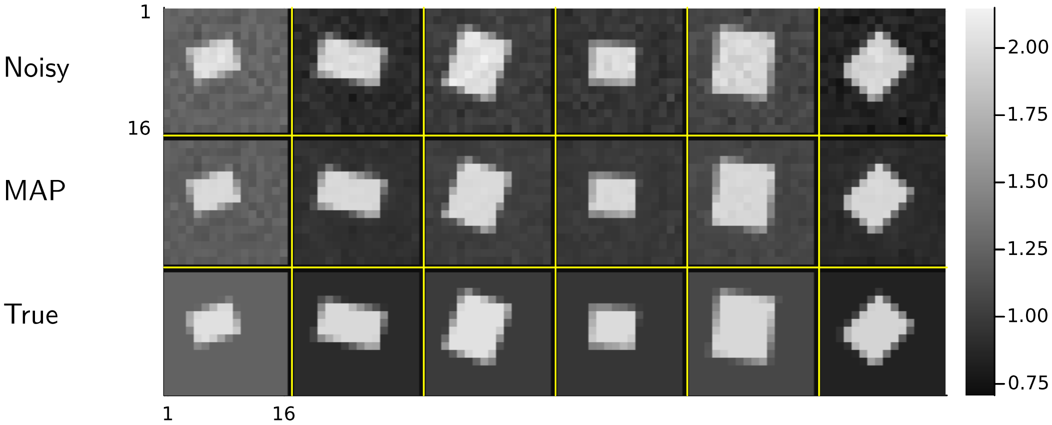
$$\mathrm{p}(\boldsymbol{x}) \propto \mathrm{e}^{-\beta |x_2 - x_1|}$$



Following trends in score matching [13, 26]
Adding gaussian noise to training data $\equiv$ smoothing score function

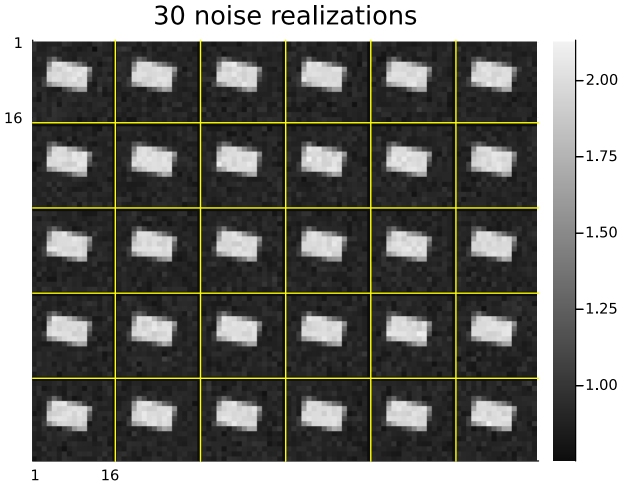


Noisy 29.5dB, MAP 29.9dB, True



Uncertainty?

- Sample from $p(\boldsymbol{x}|\boldsymbol{y})$
- Perform multiple realizations



Multiple realizations