Deep Learning in Inverse Imaging Problems

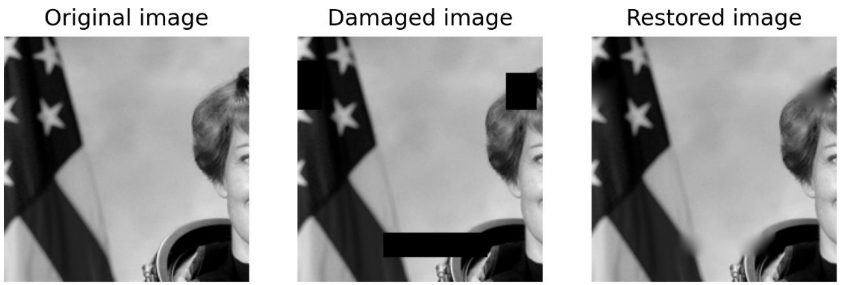
Eugenia Cao

Definitions

Inverse problems - reconstructing an unknown signal, image, or multidimensional volume from observations.

The observations are obtained from the unknown data by a **forward process**, which is typically non-invertible. ➡ ill-posed

Eg. deblurring, deconvolution, inpainting, compressed sensing, superresolution



References: https://tristanvanleeuwen.github.io/IP_and_Im_Lectures/intro.html

Mathematical Definition

Forward Model: $\mathbf{y} = f(\mathbf{x}) + \eta$

The problem of reconstructing an image x from measurements y with knowledge on the forward model f() is usually treated as a **Bayesian inference problem**:

$$\mathbf{x}^* = \arg\max_{\mathbf{x}} \underline{\log p(\mathbf{x}|\mathbf{y})} + \underline{\log p(\mathbf{x})}$$

Data-fitting term (maximum likelihood)

Prior term/Regularization (case specific)

References: https://tristanvanleeuwen.github.io/IP_and_Im_Lectures/intro.html

Image Prior

Hand-crafted priors:

- Pro: good interpretability
- Con: requires careful picking and hyperparameter tuning; may introduce human bias

Data-driven priors:

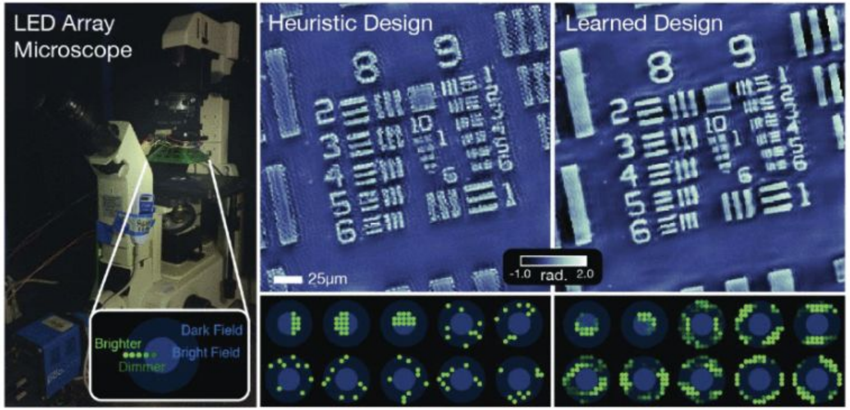
- Pro: ability to adapt to the specific characteristics of the data
- Con: requires large dataset; black box

Example: Microscopy

Challenge: **imaging artifacts & noise**

Artifacts: CCD, invisible pixel boundary (magnification), photobleaching (organic molecules), compression (posterization, contouring, blurring, blocking)

Noise: dark noise, read noise, photon shot noise



[Kellman 2019]

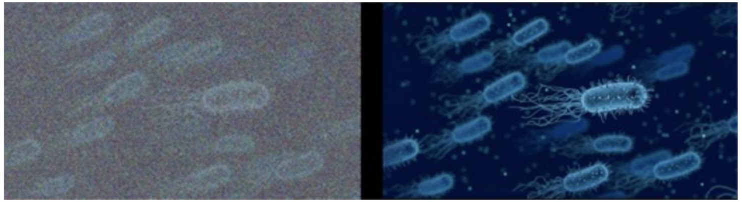


Figure 12 Images (left) before and (right) after noise correction.

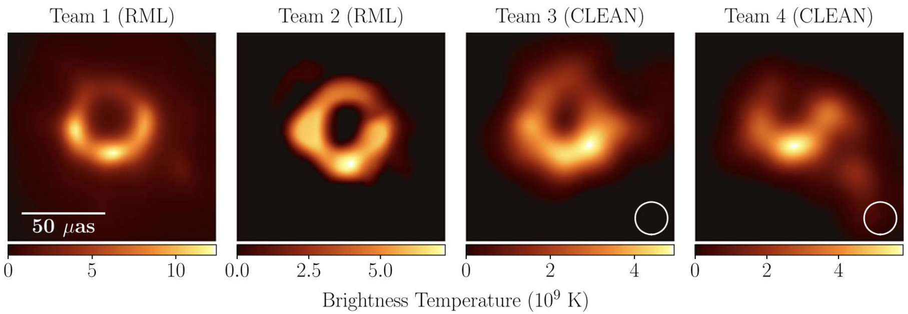
[Balaji 2024]

References: https://ieeexplore.ieee.org/abstract/document/8747339,
Details on imaging challenges: https://academic.oup.com/mt/article/32/3/22/7687037

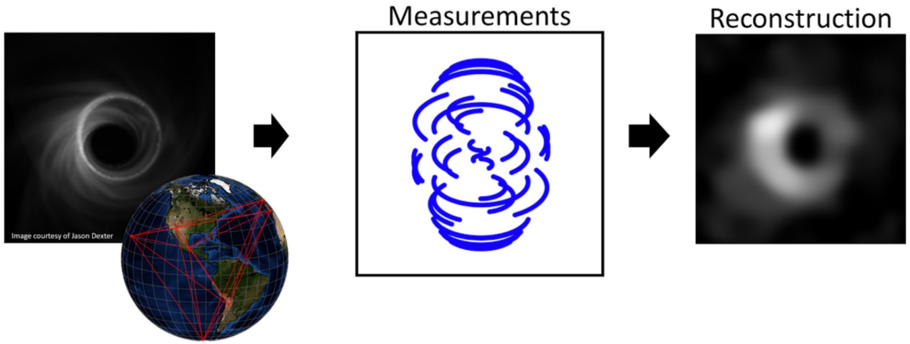
Example: Black Hole

Challenge: **limited data sources**

Using only a few points to infer the whole picture + noise + convolved Point Spread Function (instruments)



[The Event Horizon Telescope Collaboration 2019]



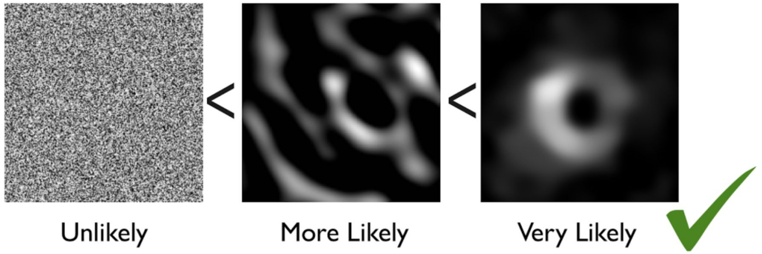
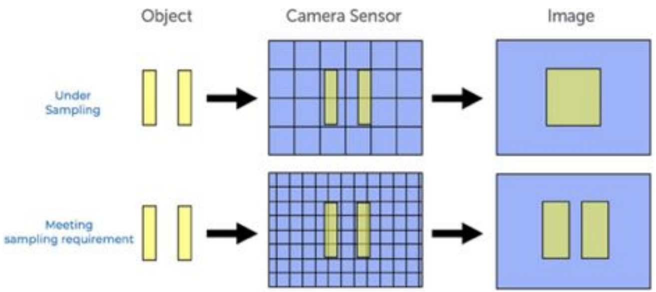


Image credit: Katie Bouman

Left square image courtesy of Jason Dexter. Image credit: Katie Bouman

References: https://www.jpl.nasa.gov/edu/resources/teachable-moment/how-scientists-captured-the-first-image-of-a-black-hole/

Example: MRI



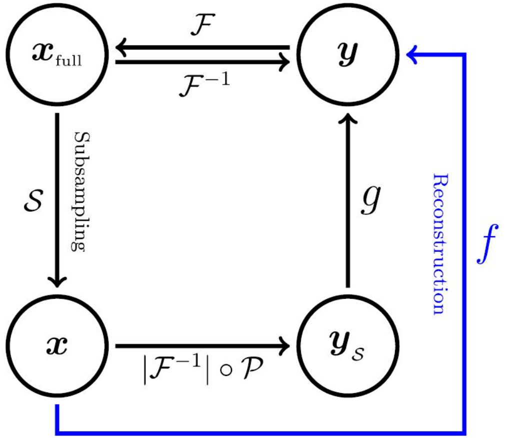
[Balaji 2024]

Challenge: **undersampled k-space**

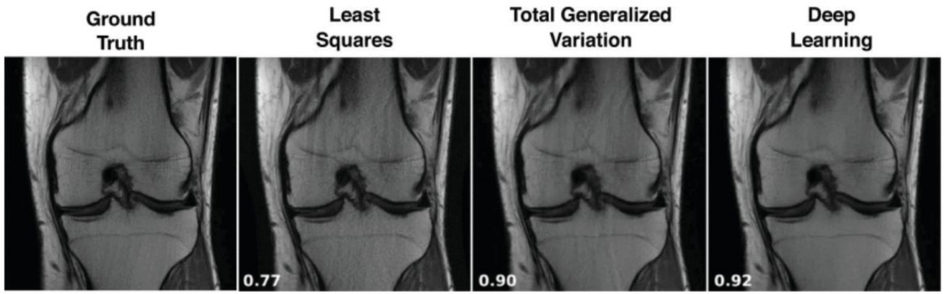
Scan time is roughly proportional to the number of time-consuming phase-encoding steps in k-space

Compressed sensing MRI: uses prior information on MR images of the unmeasured k-space data to eliminate or reduce aliasing artifacts

Parallel MRI: installs multiple receiver coils and uses space-dependent properties of receiver coils to reduce aliasing artifacts



[Hyun 2018]



[Knoll 2020]

References: https://ieeexplore.ieee.org/abstract/document/8962951

Supervised vs Unsupervised

Supervised:

Use a **matched dataset** of ground truth images x and corresponding measurements y, done by **simulating/implementing** the forward operator on clean data.

Con: sensitive to changes or uncertainty to the forward operator

Unsupervised:

Do not rely on a matched dataset of images x and measurements y.

3 types: (1) use **unpaired** ground truth images and measurements, (2) leverage ground truth images only, and (3) use only measurements.

Supervised Models

Neural Fields

Deep Image Prior - Untrained Networks

Given measurements **y** and the forward operator **A**, DIP initializes a generative network **Gθ : Rk→Rn** with a fixed random input vector **z∈R^k** , and optimizes over the network weights **θ** .

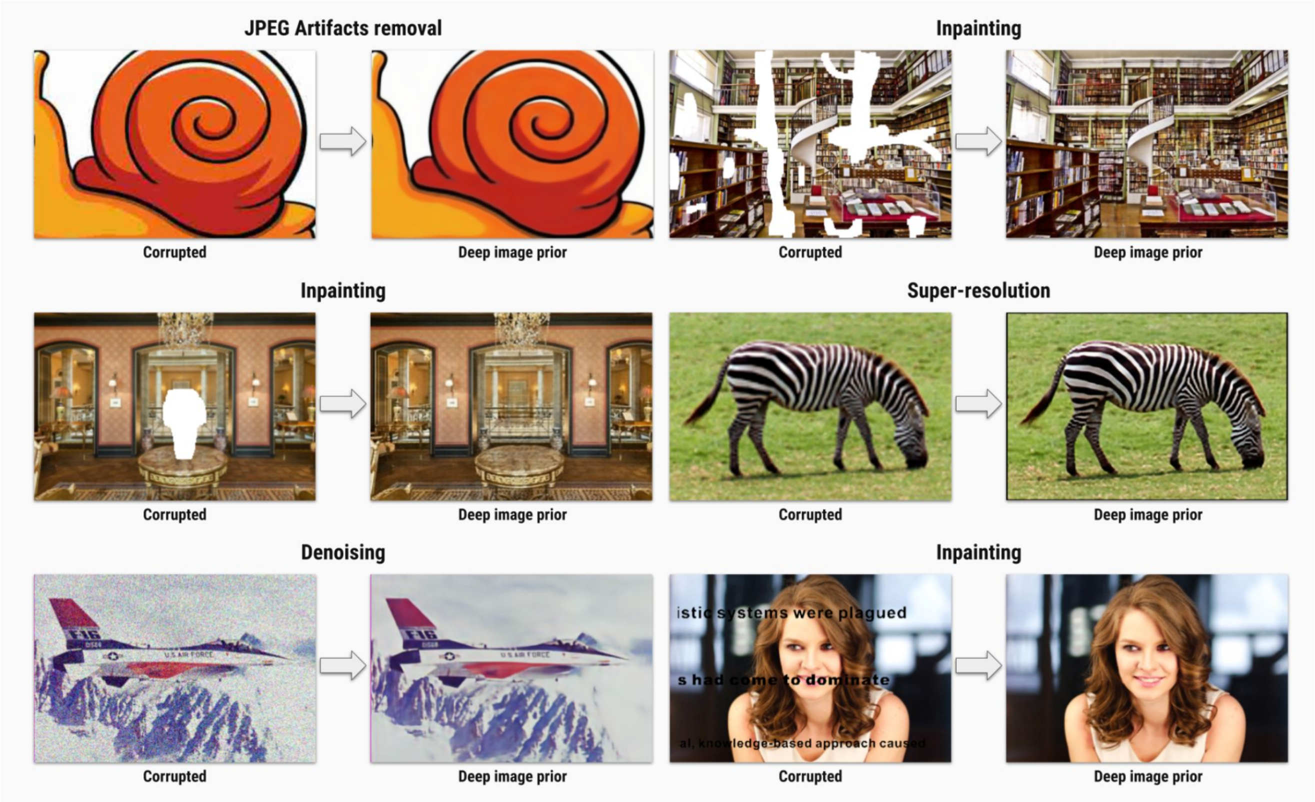
$$\theta^* = \arg\min_{\theta} \|\mathcal{A}(G_\theta(\boldsymbol{z})) - \boldsymbol{y}\|^2.$$

Implemented **early stopping** to prevent fitting the noise. (gradient descent must be stopped before it converges)

Explanation: CNN are biased towards smooth, “natural” images, and hence smooth components of an image will be reconstructed before the noisier components in the measurements.

Extension: Deep Decoder

Deep Image Prior - Untrained Networks



[Ulyanov 2018]

References: https://dmitryulyanov.github.io/deep_image_prior

Uncertainty Quantification

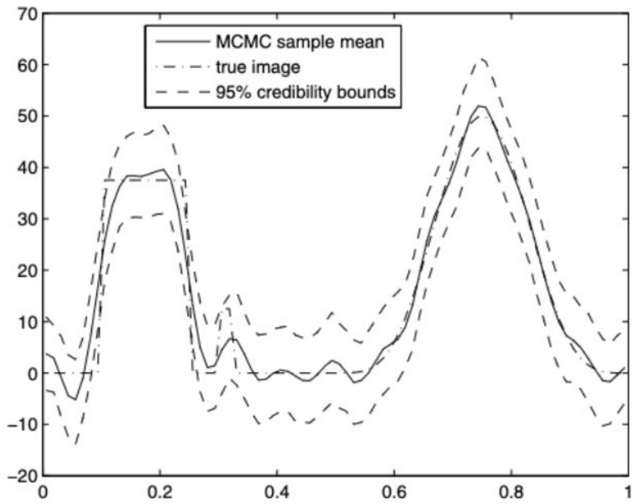
Sources of uncertainty:

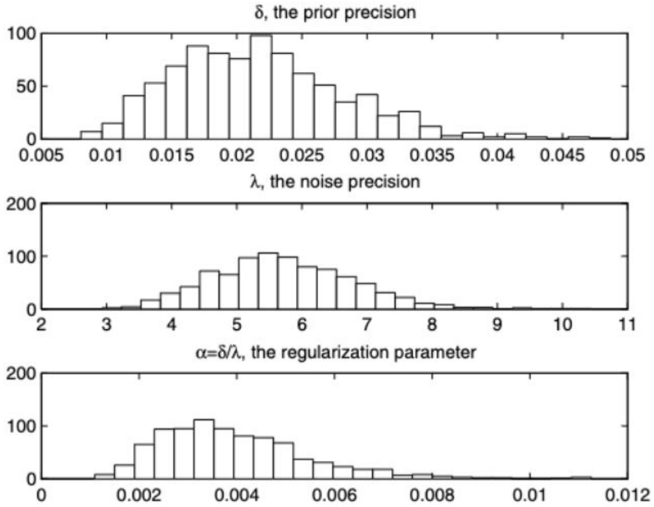
- **Aleatoric**: inherent uncertainty on the target image for a given measurement vector due to the ill-posed nature of the inverse problem
- **Epistemic**: uncertainty on the parameters of a neural network by putting a probability distribution on the parameters and computing the posterior distribution of the parameters given a training dataset

MCMC (Markov chain Monte Carlo) Sampling (2012)

Approximate the **posterior distribution** of a hidden image via sampling.

Prohibitively slow for high dimensional inverse problems.





[Bardsley, J. M. 2012]

References: https://epubs.siam.org/doi/epdf/10.1137/11085760X

Variational Bayesian Methods

Instead of directly computing the exact posterior distribution, Variational Bayes (VB) inference approximates the posterior distribution using **a family of trial probability distributions q(m)** of known form.

The hyperparameters θ for the trial distribution are determined by minimizing the **KL divergence** between the trial and posterior distributions as

$$q^*(m) = \underset{\theta}{\operatorname{argmin}}\, \mathrm{KL}(q(m)|p(m|d)) \text{ , where } q^*(m) \text{ is the trial distribution}$$

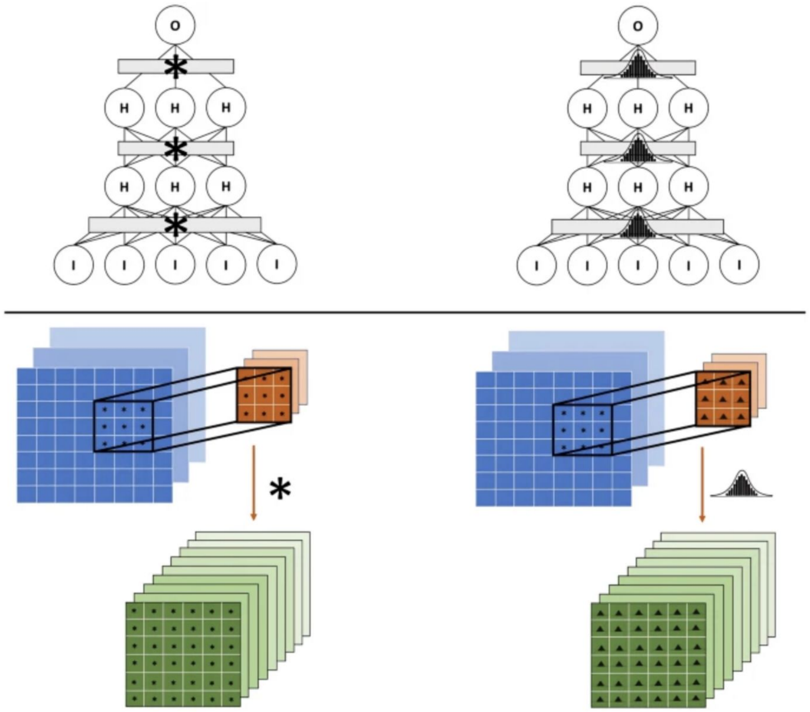
Much **faster** than sampling methods (e.g., MCMC), and typically achieve comparable performance.

References: https://link.springer.com/article/10.1007/s10462-011-9236-8

Bayesian Neural Networks

Compare to regular deep neural network, the weights of a Bayesian network are modeled as **probability distributions**, so that **different predictions are obtained** every time the network is executed.

Can be trained with variational inference through back-propagation using Bayes-by-backprop with the local reparametrization trick.



Deterministic (left) vs bayesian (right) for fully connected layers (up) and convolutional layers (down)

[Bargagna 2023]

Empirical Sampling

Obtaining samples by solving a regularized inverse problem multiple times with different choices of regularizer hyper-parameters and image initializations.

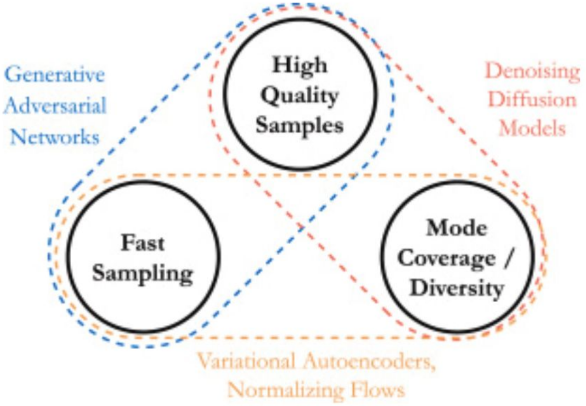
Cautions! Although the mean and standard deviation of these images provide a measure of uncertainty, there is no expectation that these samples satisfy a posterior distribution defined by the measurement data. -> In fact, quantifies the reconstruction uncertainty due to choices in the reconstruction methods (eg. regularizer hyper-parameters), instead of the real uncertainty.

Generative Models

GAN = generator (capture the distribution of true samples) **+ critic/discriminator** (estimates the probability of a given sample coming from the real dataset). Cons: unstable (mode collapse, vanishing gradients & convergence)

VAE: optimize by maximizing the **evidence lower bound** (ELBO). Cons: balancing issue, variable collapse phenomenon

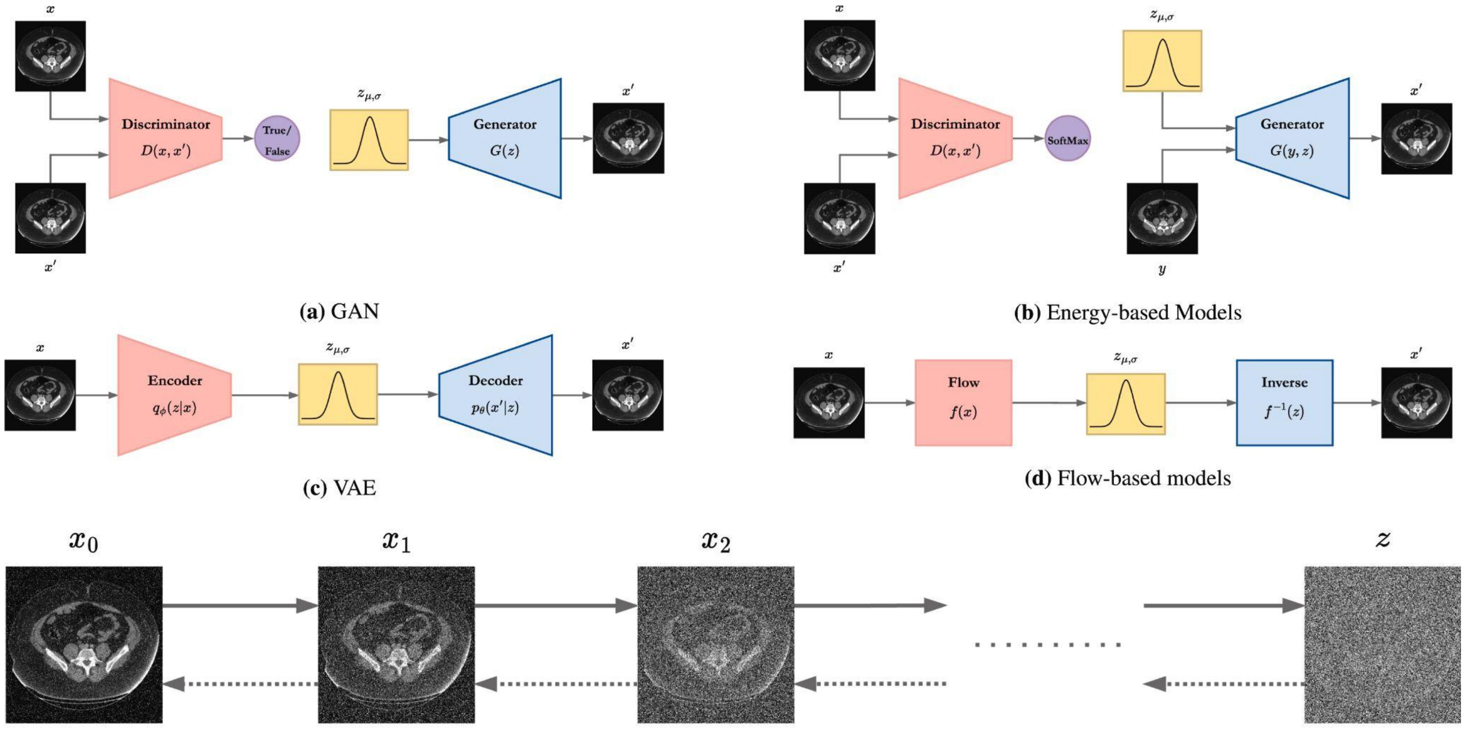
Normalizing Flows: explicitly learn the data distribution - transforms a simple distribution by applying a sequence of invertible transformation functions



[Kazerouni 2023]

References: https://www.sciencedirect.com/science/article/pii/S1361841523001068

Generative Models



(a) GAN

(b) Energy-based Models

(c) VAE

(d) Flow-based models

(e) Diffusion Models

[Kazerouni 2023]

References: https://www.sciencedirect.com/science/article/pii/S1361841523001068

Diffusion Models - Variational Perspective

Models that use **variational inference** to approximate the target distribution, generally by minimizing the Kullback–Leibler (KL) divergence between the approximate and target distributions.

Denoising Diffusion Probabilistic Models (DDPMs)

- defines the forward diffusion process as a Markov Chain where Gaussian noise is added in successive steps to obtain a set of noisy samples.

References: https://www.sciencedirect.com/science/article/pii/S1361841523001068

Diffusion Models - Score Perspective

Models rely on a **maximum likelihood-based** estimation approach, using the score function of the log-likelihood of the data to estimate the parameters of the diffusion process.

Noise conditioned score networks (NCSNs)

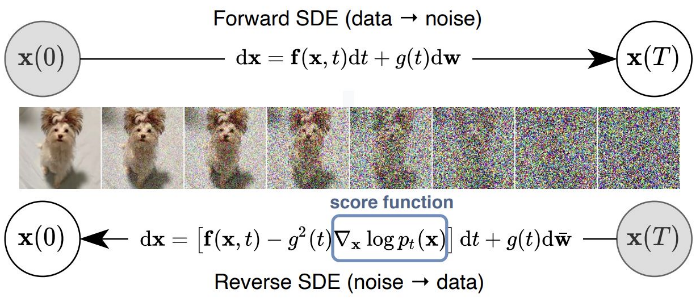
- focus on estimating the derivative of the log density function of the perturbed data distribution at different noise levels
- Trained a score network to approximate the score of p(x)

$$s_\theta(x) \approx \nabla_x \log p(x) \qquad \mathbb{E}_{x \sim p(x)} \| s_\theta(x) - \nabla_x \log p(x) \|_2^2 .$$

Hard to calculate log density -> not scalable

References: https://www.sciencedirect.com/science/article/pii/S1361841523001068

Diffusion Models - Score Perspective



Stochastic Differential Equations (SDEs)

[Song 2021]

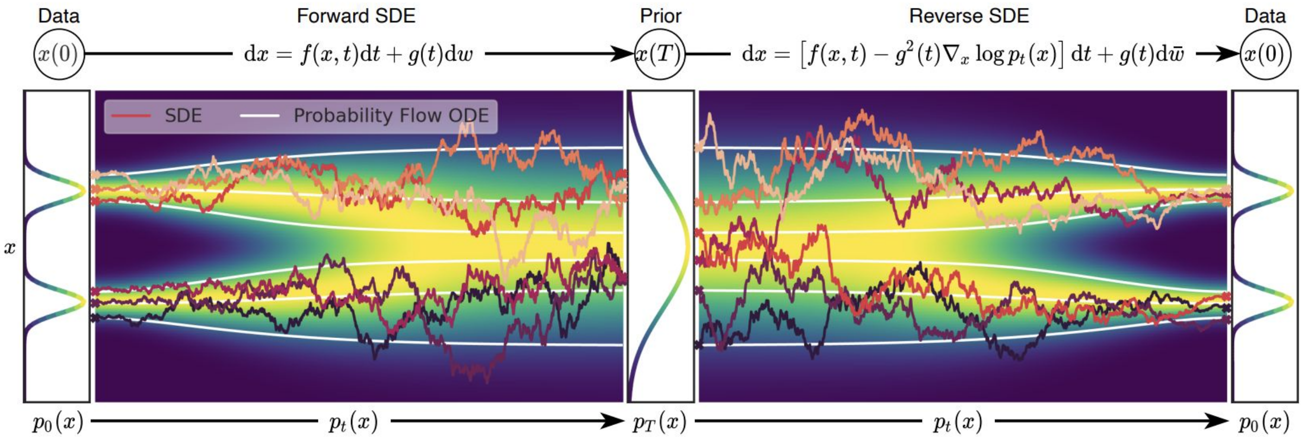


Figure 2: **Overview of score-based generative modeling through SDEs**. We can map data to a noise distribution (the prior) with an SDE (Section 3.1), and reverse this SDE for generative modeling (Section 3.2). We can also reverse the associated probability flow ODE (Section 4.3), which yields a deterministic process that samples from the same distribution as the SDE. Both the reverse-time SDE and probability flow ODE can be obtained by estimating the score $\nabla_{\mathbf{x}} \log p_t(\mathbf{x})$ (Section 3.3).

References: https://www.sciencedirect.com/science/article/pii/S1361841523001068

More on Diffusion Models

Tutorial on Diffusion Models for Imaging and Vision

Stanley Chan[1]

January 9, 2025

Abstract. The astonishing growth of generative tools in recent years has empowered many exciting applications in text-to-image generation and text-to-video generation. The underlying principle behind these generative tools is the concept of *diffusion*, a particular sampling mechanism that has overcome some longstanding shortcomings in previous approaches. The goal of this tutorial is to discuss the essential ideas underlying these diffusion models. The target audience of this tutorial includes undergraduate and graduate students who are interested in doing research on diffusion models or applying these tools to solve other problems.

Caveats/Beware/Failures

Robustness to Different Forward Model at Test Time than at Train

- learning to reconstruct MRI images for a scanner at one clinic and then attempting to use that learned algorithm to reconstruct MRI images for a (subtly different) scanner at another clinic.

Recovering Features Not Represented by Training Data

- Training set is not necessity a holistic representation of testing set.

Difficult to Interpret

- Eg: deep image prior & supervised models.

Creation of Artifacts

- Can generate realistic looking images, even when the features in the image are not actually present.

Failure Modes May be Hard to Recognize

- What if the model always returns a high quality image

Other Useful Review Articles

https://lukeli0425.github.io/Course_Projects/A_Survey_on_Black_Hole_Image_Reconstruction.pdf

https://tristanvanleeuwen.github.io/IP_and_Im_Lectures/intro.html

https://ieeexplore.ieee.org/abstract/document/9084378?figureId=fig2#fig2

https://link.springer.com/content/pdf/10.1007/978-0-387-69277-7.pdf