UPPSALA UNIVERSITET

Denoising Diffusion Probabilistic Model (DDPM)

Li Ju, Usama Zafar

- Aim: We want to generate images!
- We actually mean: To sample from $P(\boldsymbol{X}_0), \boldsymbol{X}_0 : \mathbb{R}^W \times \mathbb{R}^H \times \mathbb{R}^3$ (fixed sized image).
- Not feasible to directly model $P(\boldsymbol{X}_0)$ (high dimension).

Images generate by Denoising Diffusion Probablistic Model (DDPM)[1]

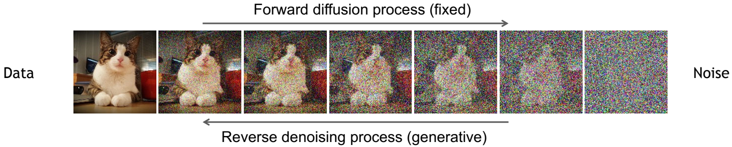
[1] Ho et al. [2020]

General Idea of DDPM

"Learn to Construct by Destroying"

Denoising diffusion models consist of two processes:

1. *Forward diffusion*: Gradually add noise to original image x_0.
2. *Reverse denoising*: Learn (approximate) the process of denoising.



Forward Diffusion Process

The forward process is defined as a *fixed* Markov Chain:

Time Step $0 \to T$

$$\boldsymbol{X}_0 \longrightarrow \boldsymbol{X}_1 \longrightarrow \cdots \longrightarrow \boldsymbol{X}_T$$

$$\boldsymbol{X}_t | \boldsymbol{x}_{t-1} \sim \mathcal{N}(\boldsymbol{X}_t; \sqrt{1-\beta_t}\boldsymbol{x}_{t-1}, \beta_t \boldsymbol{I})$$

where $\beta_t \in (0, 1),\ \forall t \in \{1, \ldots, T\}$ is a predefined sequence of monotonically increasing parameters.

Forward Diffusion Process

Let $\overline{\alpha}_\tau = \prod_{t=1}^{\tau}(1-\beta_t)$. Any any time step we have

$$\boldsymbol{X}_t|\boldsymbol{x}_0 \sim \mathcal{N}(\boldsymbol{X}_t; \sqrt{\overline{\alpha}_t}\boldsymbol{x}_0, (1-\overline{\alpha}_t)\boldsymbol{I})$$

Due to the choice of β_t values such that $\overline{\alpha}_T \to 0$, further we have

$$\boldsymbol{X}_T|\boldsymbol{x}_0 \sim \mathcal{N}(\boldsymbol{X}_T; \boldsymbol{0}, \boldsymbol{I}) \text{ and } \boldsymbol{X}_T \sim \mathcal{N}(\boldsymbol{X}_T; \boldsymbol{0}, \boldsymbol{I})$$

approximately.

Forward Diffusion Process

Let $\overline{\alpha}_\tau = \prod_{t=1}^{\tau} (1 - \beta_t)$. Any any time step we have

$$\boldsymbol{X}_t | \boldsymbol{x}_0 \sim \mathcal{N}(\boldsymbol{X}_t; \sqrt{\overline{\alpha}_t}\boldsymbol{x}_0, (1 - \overline{\alpha}_t)\boldsymbol{I})$$

Due to the choice of β_t values such that $\overline{\alpha}_T \to 0$, further we have

$$\boldsymbol{X}_T | \boldsymbol{x}_0 \sim \mathcal{N}(\boldsymbol{X}_T; \boldsymbol{0}, \boldsymbol{I}) \text{ and } \boldsymbol{X}_T \sim \mathcal{N}(\boldsymbol{X}_T; \boldsymbol{0}, \boldsymbol{I})$$

approximately.

To summarize,

- We know $p(\boldsymbol{X}_t | \boldsymbol{x}_0), \forall t \in [1, \ldots, T]$ and $p(\boldsymbol{X}_T)$
- We don't know $p(\boldsymbol{X}_0)$, which is intractable from $p(\boldsymbol{X}_{0:T})$.

Reverse Denoising Process

Variational Family

We can approximate it with $Q_\theta \in \mathcal{Q}_\theta$ from a parameterised variational family which is a reverse Markov chain:

Time Step $T \to 0$

$$\boldsymbol{X}_0 \leftarrow \boldsymbol{X}_1 \leftarrow \cdots \leftarrow \boldsymbol{X}_T$$

i.e.

$$Q_\theta(\boldsymbol{X}_{0:T}) = Q_\theta(\boldsymbol{X}_T) \prod_{t=1}^{T} Q_\theta(\boldsymbol{X}_{t-1} | \boldsymbol{x}_t)$$

Further we assume $P(\boldsymbol{X}_T) = Q_\theta(\boldsymbol{X}_T) = \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$.

Reverse Denoising Process

Derivation of Loss: I

We want to minimise $\mathrm{KL}\left(P(\boldsymbol{X}_{0:T}) \| Q_\theta(\boldsymbol{X}_{0:T})\right)$ (Why?)

$$
\begin{aligned}
&\mathrm{KL}\left(P(\boldsymbol{X}_{0:T}) \| Q_\theta(\boldsymbol{X}_{0:T})\right) \\
&= \mathbb{E}_P\left[\ln \frac{p(\boldsymbol{X}_{0:T})}{q_\theta(\boldsymbol{X}_{0:T})}\right] \\
&= \mathbb{E}_P\left[\ln \frac{p(\boldsymbol{X}_{1:T} | \boldsymbol{x}_0)}{q_\theta(\boldsymbol{X}_{1:T} | \boldsymbol{x}_0)} - \ln q_\theta(\boldsymbol{X}_0) + \ln p(\boldsymbol{X}_0)\right] \\
&= \mathbb{E}_{P(\boldsymbol{X}_0)}\left[\underbrace{\mathbb{E}_{P(\boldsymbol{X}_{1:T} | \boldsymbol{x}_0)}\left[\ln \frac{p(\boldsymbol{X}_{1:T} | \boldsymbol{x}_0)}{q_\theta(\boldsymbol{X}_{1:T} | \boldsymbol{x}_0)}\right] - \ln q_\theta(\boldsymbol{X}_0)}_{\text{negative ELBO}} + \ln p(\boldsymbol{X}_0)\right]
\end{aligned}
$$

Reverse Denoising Process

Derivation of Loss: II

We further factorise $p(\boldsymbol{X}_{1:T}|\boldsymbol{x}_0)$ and $q_\theta(\boldsymbol{X}_{1:T}|\boldsymbol{x}_0)$:

$$
\begin{aligned}
&\mathbb{E}_{P(\boldsymbol{X}_{1:T}|\boldsymbol{x}_0)}\left[\ln \frac{p(\boldsymbol{X}_{1:T}|\boldsymbol{x}_0)}{q_\theta(\boldsymbol{X}_{1:T}|\boldsymbol{x}_0)}\right] - \ln q_\theta(\boldsymbol{X}_0) \\
=\ &\underbrace{-\mathbb{E}_{P(\boldsymbol{X}_1|\boldsymbol{x}_0)}\left[\ln q_\theta(\boldsymbol{X}_0|\boldsymbol{x}_1)\right]}_{L_0} + \\
&\underbrace{\sum_{t=2}^{T} \mathbb{E}_{P(\boldsymbol{X}_t|\boldsymbol{x}_0)}\left[\mathrm{KL}(P(\boldsymbol{X}_{t-1}|\boldsymbol{x}_t, \boldsymbol{x}_0) \| Q_\theta(\boldsymbol{X}_{t-1}|\boldsymbol{x}_t))\right]}_{L_{t-1}} + \\
&\underbrace{\mathrm{KL}(P(\boldsymbol{X}_T|\boldsymbol{x}_0) \| Q_\theta(\boldsymbol{X}_T))}_{L_T \text{ (const.)}}
\end{aligned}
$$

Reverse Denoising Process

Parameterisation

UPPSALA UNIVERSITET

For L_{t-1}, we assume $Q_\theta(\boldsymbol{X}_{t-1}|\boldsymbol{x}_t)$ are Gaussian distributions for all $t \in [0, T]$:

$$Q_\theta(\boldsymbol{X}_{t-1}|\boldsymbol{x}_t) = \mathcal{N}\left(\boldsymbol{\mu}_\theta(\boldsymbol{x}_t, t), \beta_t \boldsymbol{I}\right)$$

We can use a neural network $f_\theta(\boldsymbol{x}_t, t)$ to approximate $\boldsymbol{\mu}_\theta(\boldsymbol{x}_t, t)$.

For L_0, similarly, we use another neural network $g_\theta(\boldsymbol{x}_1)$ to approximate $\ln q_\theta(\boldsymbol{X}_0|\boldsymbol{x}_1)$.

Summary

- DDPMs model complex probability distribution by:
  - A forward process that transforms it into a simple distribution by gradual distortion.
  - A reverse process as the variational family to approximate the posterior, thanks to Bayesian inference.
- DDPM is deeply rooted on Bayesian statistics, utilising variational inference and neural networks.
- Implementation details are skipped (residual modelling, reparameterisation trick, etc.).

References

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *CoRR*, abs/2006.11239, 2020. URL `https://arxiv.org/abs/2006.11239`.

UPPSALA
UNIVERSITET

Thank you for your attention!