دیفیوژن

taha hoseinpour

tahahoseinpourasli@gmail.com

مقدمه

دیفیوژن چگونه کار میکند؟

به صورت شهودی میتوان گفت که فرایند دیوفیوژن به این صورت است که ما تصویر را با استفاده از نویزهای گاوسی به صورت iterative در تعداد n مرحله در یک زنجیره مارکو به عکس وارد میکنیم تا در پایان این n مرحله تصویر مورد نظر ما به یک تصویر کاملاً نویزی میل بکند.

توضیح تفصیلی فرایند: در واقع این فرایند شامل دو مرحله اصلی است:

۱. **فرایند Forward (نویزینگ):** در این مرحله ما به تدریج و در گام‌های کوچک نویز گاوسی به تصویر اصلی اضافه می‌کنیم تا در نهایت به یک نویز کامل گاوسی برسیم.
۲. **فرایند Reverse (دینویزینگ):** در این مرحله ما یاد می‌گیریم که چطور این فرایند را برعکس کنیم و از نویز کامل به تصویر معنادار برگردیم.

پس حالا ایده اصلی این مدل در اینجا هست که ما می‌خواهیم دینویزری را طراحی بکنیم که بتواند در هر مرحله از زمان یعنی t بتواند تخمین بزند که چقدر باید تصویر را دینویز کند و چه تغییری باید در تصویر نویزی کامل ایجاد کند تا ما بتوانیم برگردیم و این نویزها را به درستی شناسایی کند. پس حالا اینگونه میتوان از یک استیت نویزی کامل گاوسی یعنی نویز کامل به سمت یک تصویر معنادار حرکت کرد.

اهمیت شبکه عصبی دینویزر: شبکه عصبی دینویزر که معمولاً یک U-Net است، در هر گام زمانی t، تصویر نویزی x_t و زمان t را به عنوان ورودی دریافت می‌کند و نویز موجود در تصویر را تخمین می‌زند. این تخمین نویز سپس از تصویر نویزی کم می‌شود تا تصویر کمتر نویزی به دست آید.

در اینجا باید از manifold hypothesis هم صحبت کرد که فرضی هست که دیتاست ما در یک فضای منیفلدی که تعداد ابعادش بالا هستند (میتوانید یک فضای اقلیدسی R^n را فرض کنید) معمولاً این دیتاهای ما در یک فضای منیفلدی با تعداد ابعاد کمتر در همان منیفلد تجمیع می‌شوند. پس ما هم از این فرض استفاده می‌کنیم. در واقع فرایند دیفیوژن اینگونه هست که ما در فضای منیفلدی داریم جستجو و سرچ انجام می‌دهیم تا به این فضای منیفلدی با بعد کمتر خاص درونش نزدیک‌تر بشویم.

توضیح Manifold Hypothesis : فرض منیفلد بیان می‌کند که داده‌های واقعی (مثل تصاویر طبیعی) اگرچه در فضای با ابعاد بالا (مثل $R^{256\times256\times3}$ برای تصاویر رنگی) قرار دارند، اما در واقع روی یک منیفلد با ابعاد بسیار کمتر متمرکز هستند. به عبارت دیگر، تمام تصاویر ممکن در فضای پیکسل‌ها معنادار نیستند،

بلکه فقط زیرمجموعه کوچکی از آنها تصاویر واقعی و معنادار هستند.

به قول آقای Song یکی از نویسندگان مقاله دیفیوژن مدل:

Creating noise from data is easy; creating data from noise is generative modeling

فرایند Diffusion Forward

فرض کنید که یک data یا عکس از distribution واقعی data سمپل کردیم یعنی $x_0 \sim q(x)$.

توضیح نماد: x_0 نمایانگر تصویر اصلی (بدون نویز) است که از توزیع واقعی داده‌ها $q(x)$ نمونه‌برداری شده است. این $q(x)$ توزیع ناشناخته‌ای است که تمام تصاویر واقعی را در بر می‌گیرد.

حالا در T مرحله یعنی مراحل $1, 2, 3, ..., T$ ما بهش نویز اضافه یا inject می‌کنیم تا دنباله‌ای از تصویر نویزی را داشته باشیم به این طریق که: $x_1 \to x_2 \to ... \to x_T$

احتمال اینکه از تصویر x_{t-1} به تصویر x_t برویم احتمال توزیع گاوسی چندمتغیره می‌باشد:

$$q(x_t \mid x_{t-1}) = \mathcal{N}(x_t; \sqrt{1-\beta_t}x_{t-1}, \beta_t I)$$

توضیح این رابطه:

- β_t در واقع size step ما یا طول گام ما هست یعنی اینکه در طول هر گام چقدر می‌خواهیم که نویزی را به تصویر وارد بنماییم
- $\sqrt{1-\beta_t}$ ضریب تضعیف سیگنال اصلی است. هرچه β_t بزرگ‌تر باشد، سیگنال اصلی بیشتر تضعیف می‌شود
- $\beta_t I$ ماتریس کوواریانس نویز اضافه شده است، که I ماتریس همانی است
- این فرمول بیان می‌کند که x_t توزیع نرمال دارد با میانگین $\sqrt{1-\beta_t}x_{t-1}$ و واریانس β_t

و همینطور جوری این فرایندها را در هر مرحله تعریف کرده‌ایم که نویزی که برای مثال در مرحله i ام وارد می‌شود با نویزی که در مرحله j ام وارد می‌شود مستقل می‌باشد پس فرایندهای ما مستقل هستند پس داریم:

$q(x_{1:T} \mid x_0) = \prod_{t=1}^{T} q(x_t|x_{t-1})$

توضیح فرض مارکو: زیرا طبق فرض قاعده زنجیره‌ای احتمالات و فرض مستقل بودن فرایندها از هم یا فرض زنجیره مارکوف داریم:

$q(x_{1:T} \mid x_0) = q(x_1 \mid x_0)\, q(x_2 \mid x_0, x_1) \cdots q(x_T \mid x_0, x_1, \ldots, x_{T-1})$

and by Markov property assumption:

$q(x_t \mid x_{0:t-1}) = q(x_t \mid x_{t-1})$

معنی فرض مارکو: فرض مارکو بیان می‌کند که آینده فقط به حال وابسته است، نه به گذشته. به عبارت دیگر، اگر می‌دانیم x_{t-1} چیست، دانستن $x_{t-2}, x_{t-3}, \ldots$ اطلاعات اضافی برای پیش‌بینی x_t نمی‌دهد.

که در آن $x_{1:T}$ یعنی احتمال پیشامد اینکه تصویر نویزی اول x_1 باشد و تصویر نویزی دوم x_2 باشد و به همین ترتیب...

همچنین می‌توان فرایند نویزی که داشتیم به این صورت بود را:

$q(x_{1:T} \mid x_0) = \prod_{t=1}^{T} q(x_t|x_{t-1})$

با فرض اینکه $\alpha_t = 1 - \beta_t$ و $\bar{\alpha_t} = \prod_{i=1}^{t} \alpha_i$ را نیز به این صورت با استفاده از نویز گاوسی که از توزیع استاندارد نرمال پیروی می‌کند هم نشان بدهیم:

توضیح نمادگذاری جدید:

- $\alpha_t = 1 - \beta_t$: این تبدیل برای سادگی محاسبات انجام می‌شود
- $\bar{\alpha_t} = \prod_{i=1}^{t} \alpha_i = \alpha_1 \times \alpha_2 \times \ldots \times \alpha_t$: این حاصل‌ضرب تجمعی است که نشان می‌دهد پس از t گام چه مقدار از سیگنال اصلی باقی مانده

$$\mathbf{x}_t = \sqrt{\alpha_t}\mathbf{x}_{t-1} + \sqrt{1-\alpha_t}\boldsymbol{\epsilon}_{t-1} \quad \text{where؛ } \boldsymbol{\epsilon}_{t-1} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$$

توضیح این فرمول: این فرمول نشان می‌دهد که چطور می‌توانیم یک گام نویزینگ را به صورت جبری نمایش دهیم:

- $\sqrt{\alpha_t}\mathbf{x}_{t-1}$: بخشی از تصویر قبلی که حفظ می‌شود
- $\sqrt{1-\alpha_t}\boldsymbol{\epsilon}_{t-1}$: نویز گاوسی که اضافه می‌شود
- $\boldsymbol{\epsilon}_{t-1} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$: نویز از توزیع نرمال استاندارد

زیرا توزیع گاوسی استاندارد را می‌توان با ضرب و جمع کردن دو عدد مورد نظر خودمان به توزیع گاوسی دلخواهی که نویز را وارد می‌کرد بدل بکنیم.

ویژگی مهم توزیع گاوسی: اگر $Z \sim \mathcal{N}(0,1)$ (توزیع نرمال استاندارد) باشد، آنگاه $aZ + b \sim \mathcal{N}(b, a^2)$. این ویژگی به ما اجازه می‌دهد هر توزیع گاوسی را از توزیع استاندارد بسازیم.

اثبات جمع دو توزیع گاوسی

فرض کنیم که دو توزیع گاوسی داریم به این صورت $f(z) \sim \mathcal{N}(\mu_1, \sigma_1^2)$ و $g(z) \sim \mathcal{N}(\mu_2, \sigma_2^2)$ و همچنین فرض می‌کنیم که این دو توزیع از یکدیگر مستقل هستند آنگاه توزیع مجموع این دو توزیع چه می‌شود؟

اهمیت این اثبات: این اثبات برای درک اینکه چرا می‌توانیم چندین گام نویزینگ را در یک گام خلاصه کنیم، بسیار مهم است.

این را برای توزیع نرمال یک متغیره نشان می‌دهیم و اثبات برای توزیع‌های چندمتغیر نیز یکسان هست:

خواهیم داشت:

$h(z) = f(z) + g(z)$

پس اگر بخواهیم توزیع احتمال برای یک نقطه مانند x محاسبه کنیم داریم احتمال می‌شود جمع تمام احتمالاتی که جمع دو متغیر مجموعاً x بشود به معنای دیگر:

مفهوم کانولوشن: برای یافتن توزیع جمع دو متغیر تصادفی مستقل، باید کانولوشن توابع چگالی آنها را محاسبه کنیم.

$h(x) = \int_{-\infty}^{\infty} f(t) \cdot g(x-t) dt$

$$f(x) = \frac{1}{\sqrt{2\pi\sigma_1^2}} \exp\left(-\frac{(x-\mu_1)^2}{2\sigma_1^2}\right), \quad g(x) = \frac{1}{\sqrt{2\pi\sigma_2^2}} \exp\left(-\frac{(x-\mu_2)^2}{2\sigma_2^2}\right).$$

$$h(x) = (f * g)(x) = \frac{1}{2\pi\sigma_1\sigma_2} \int_{-\infty}^{\infty} \exp\left(-\frac{(x-t-\mu_1)^2}{2\sigma_1^2} - \frac{(t-\mu_2)^2}{2\sigma_2^2}\right) dt$$

مراحل حل انتگرال: برای حل این انتگرال، باید نماینده را به شکل استانداردی در بیاوریم:

$$-\frac{1}{2}\left(\frac{(t-(x-\mu_1))^2}{\sigma_1^2} + \frac{(t-\mu_2)^2}{\sigma_2^2}\right) = -\frac{1}{2}\left(A\left(t-\frac{B}{A}\right)^2 - \frac{B^2}{A} + C\right)$$

که در آن:

$$A = \frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}, \quad B = \frac{x-\mu_1}{\sigma_1^2} + \frac{\mu_2}{\sigma_2^2}.$$

محاسبه انتگرال گاوسی: انتگرال گاوسی استاندارد برابر است با:

$$\int_{-\infty}^{\infty} \exp\left(-\frac{A}{2}\left(t-\frac{B}{A}\right)^2\right) dt = \sqrt{\frac{2\pi}{A}}.$$

نتیجه نهایی: پس از انجام محاسبات کامل داریم:

$$h(x) = \frac{1}{\sqrt{2\pi(\sigma_1^2+\sigma_2^2)}} \exp\left(-\frac{(x-(\mu_1+\mu_2))^2}{2(\sigma_1^2+\sigma_2^2)}\right).$$

قانون مهم: پس جمع دو توزیع گاوسی خود یک توزیع گاوسی می‌شود با میانگین $\mu_1 + \mu_2$ و واریانس $\sigma_1^2 + \sigma_2^2$.

پس به این دلیل جمع دو توزیع گاوسی ϵ_1 و ϵ_2 خود نیز یک توزیع گاوسی هست. حالا با توجه به رابطه بالا ما میانگین و واریانس جمع توزیع گاوسی دو نویز را محاسبه می‌کنیم:

به دست اوردن فرمول نویزینگ

$$\begin{aligned}\mathbf{x}_t &= \sqrt{\alpha_t}\,\mathbf{x}_{t-1} + \sqrt{1-\alpha_t}\,\boldsymbol{\epsilon}_{t-1}\\ &= \sqrt{\alpha_t}\left(\sqrt{\alpha_{t-1}}\,\mathbf{x}_{t-2} + \sqrt{1-\alpha_{t-1}}\,\boldsymbol{\epsilon}_{t-2}\right) + \sqrt{1-\alpha_t}\,\boldsymbol{\epsilon}_{t-1}\end{aligned}$$

گام به گام باز کردن رابطه: در اینجا x_{t-1} را بر حسب x_{t-2} جایگزین کرده‌ایم:

$$= \sqrt{\alpha_t\alpha_{t-1}}\,\mathbf{x}_{t-2} + \sqrt{\alpha_t(1-\alpha_{t-1})}\,\boldsymbol{\epsilon}_{t-2} + \sqrt{1-\alpha_t}\,\boldsymbol{\epsilon}_{t-1}$$

استفاده از قانون جمع گاوسی: پس همانطور که توضیح جمع گاوسی را محاسبه کردیم، جمع دو توزیع ϵ_{t-2} با ϵ_{t-1} می‌دهد $\bar{\epsilon}_{t-2}$ یعنی داریم:

دو نویز مستقل $\sqrt{\alpha_t(1-\alpha_{t-1})}\,\boldsymbol{\epsilon}_{t-2}$ و $\sqrt{1-\alpha_t}\,\boldsymbol{\epsilon}_{t-1}$ که هر دو از $\mathcal{N}(0,I)$ می‌آیند، جمع آنها برابر است با:

$$\mathcal{N}(0,(\alpha_t(1-\alpha_{t-1})+(1-\alpha_t))I) = \mathcal{N}(0,(1-\alpha_t\alpha_{t-1})I)$$

پس طبق محاسباتی که انجام دادیم $\sqrt{\alpha_t\alpha_{t-1}}\,\mathbf{x}_{t-2}+\sqrt{\alpha_t(1-\alpha_{t-1})}\,\boldsymbol{\epsilon}_{t-2}+\sqrt{1-\alpha_t}\,\boldsymbol{\epsilon}_{t-1}$ برابر میشود با:

$$\sqrt{\alpha_t\alpha_{t-1}}\mathbf{x}_{t-2} + \sqrt{1-\alpha_t\alpha_{t-1}}\bar{\boldsymbol{\epsilon}}_{t-2} \text{ where: } \bar{\boldsymbol{\epsilon}}_{t-2}. \text{ merges two Gaussians}$$

که $\boldsymbol{\epsilon}_{t-2}$ جایی هست که دو اپسیلون با هم ترکیب میشوند

ادامه فرایند بازگشتی: حالا اگر در فرمول بالا x_{t-2} را نیز بر حسب همان توزیع گاوسی و x_{t-3} بنویسیم و همین کار را تا x_0 ادامه دهیم، خواهیم داشت:

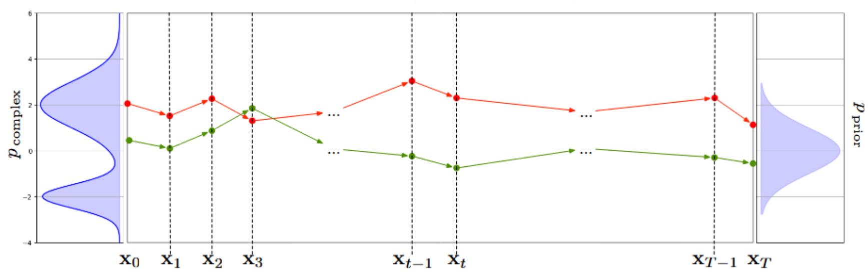
$$
\begin{aligned}
\mathbf{x}_t &= \sqrt{\alpha_t}\mathbf{x}_{t-1} + \sqrt{1-\alpha_t}\boldsymbol{\epsilon}_{t-1} && \text{where؛}\boldsymbol{\epsilon}_{t-1}, \boldsymbol{\epsilon}_{t-2}, \dots \sim \mathcal{N}(\mathbf{0}, \mathbf{I}) \\
&= \sqrt{\alpha_t \alpha_{t-1}}\mathbf{x}_{t-2} + \sqrt{1-\alpha_t \alpha_{t-1}}\bar{\boldsymbol{\epsilon}}_{t-2} && \text{where؛}\bar{\boldsymbol{\epsilon}}_{t-2}\text{Gaussians. two merges} \\
&= \dots \\
&= \sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}
\end{aligned}
$$

نتیجه نهایی:

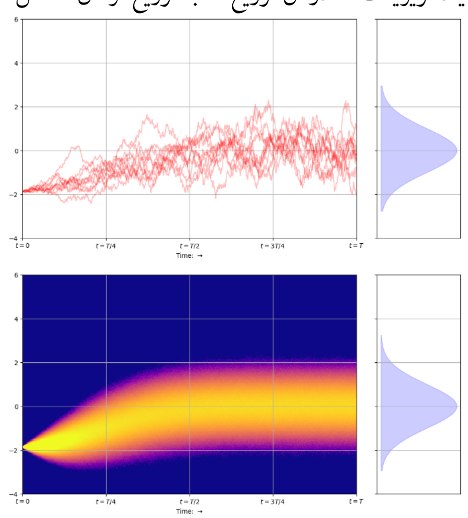
$$
q(\mathbf{x}_t|\mathbf{x}_0) = \mathcal{N}(\mathbf{x}_t; \sqrt{\bar{\alpha}_t}\mathbf{x}_0, (1-\bar{\alpha}_t)\mathbf{I})
$$

این فرمول مهم است چون به ما اجازه می‌دهد بدون اجرای t گام متوالی، مستقیماً از تصویر اصلی x_0 به تصویر نویزی در زمان t برویم. این برای آموزش مدل بسیار مفید است.

شکل ۱: روند فرایند نویزینگ که در آن توزیع ما به توزیع نرمال خالص گاوسی میل میکند



شکل ۲: روند فرایند نویزینگ که در آن توزیع ما به توزیع نرمال خالص گاوسی میل میکند



فرایند Diffusion Reverse (دینویزینگ)

حالا بعد از اینکه فرایند نویزینگ را شرح دادیم باید فرایند دینویزینگ را نیز شرح دهیم و توضیح دهیم یعنی اینکه چطوری از حالت نویزی شده می‌توان برگشت.

یعنی باید تابع $q(x_{t-1} \mid x_t)$ را محاسبه کنیم و به دست بیاوریم. معمولاً شهود ما در اینجا این است که اگر در اینجا بتاها یا طول گام نویز را به اندازه کافی کوچک در نظر بگیریم می‌توان فرض کرد که این تابع را نیز یک توزیع گاوسی فرض کرد.

چالش محاسباتی: و همچنین باید در نظر داشته باشیم که محاسبه کردن این تابع از لحاظ محاسباتی سخت و نزدیک به نشدنی است چون باید از تمام دیتاست برای محاسبه این تابع استفاده کنیم پس باید مدلی را ایجاد کنیم که بتواند این تابع را تخمین بزند.

چرا نمی‌توان به صورت مستقیم $q(x_{t-1} \mid x_t)$ را مستقیماً محاسبه کنیم؟

زیرا ما برای انجام دادن این کار باید از قانون بیز استفاده کنیم و صورت و مخرج کسر داده شده را خواهیم داشت:

$$q(x_{t-1} \mid x_t) = q(x_t \mid x_{t-1}) \frac{q(x_{t-1})}{q(x_t)}$$

و خب

این کسر برابر است با:

$$q(\mathbf{x}_{t-1}|\mathbf{x}_t) = q(\mathbf{x}_t|\mathbf{x}_{t-1}) \frac{\int q(\mathbf{x}_{t-1}|\mathbf{x}_0) q(\mathbf{x_0}) dx}{\int q(\mathbf{x}_t|\mathbf{x}_0) q(\mathbf{x_0})}$$

مشکل اصلی: و خب مشکل در این هست که ما توزیع $q(x_0)$ را به طور مستقیم نداریم و برای محاسبه این توزیع باید به کل دیتاست دسترسی داشته باشیم که همچین کاری در اکثر مواقع ممکن نیست و یا از لحاظ محاساباتی سخت میباشد پس حالا که نمیتوانیم اون توزیع را به طور مستقیم محاسبه کنیم باید مدل شبکه عصبی را طراحی کنیم که بتواند این توزیع معکوس را با دقت خوبی تخمین بزند یعنی باید $p_\theta(x_{t-1}|x_t)$ را به جوری به دست بیاریم و بهینه سازی اش کنیم که به توزیع واقعی نزدیک بشود.

توزیع $q(x_{t-1}|x_t, x_0)$:

بر خلاف توزیع $q(x_{t-1} \mid x_t)$ فرم انالتیک توزیع $q(x_{t-1}|x_t, x_0)$ را میتوان محاسبه کرد پس بگذارید اول این توزیع $q(x_{t-1}|x_t, x_0)$ را از لحاظ جبری به صورت بیشتری بسط بدهیم تا بتوان از ویژگی هایش برای انالیز بیشتر و طراحی کردن بهتر فرایند استفاده کرد.

ما تصویر اولیه و اصلی را هم در شرط اضافه میکنیم و خواهیم داشت طبق قانون بیز برای این توزیع

$$q(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_0) = q(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{x}_0)\frac{q(\mathbf{x}_{t-1}|\mathbf{x}_0)}{q(\mathbf{x}_t|\mathbf{x}_0)}$$

استفاده از فرض مارکوف: طبق فرض مارکوف داریم:

$$q(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{x}_0) = q(\mathbf{x}_t|\mathbf{x}_{t-1})$$

توضیح: چون x_t فقط به x_{t-1} وابسته است، دانستن x_0 اطلاعات اضافی نمی‌دهد.

و چون هر سه توزیع گاوسی هستند، می‌توانیم محاسبه دقیق انجام دهیم:

$$q(x_{t-1} \mid x_t, x_0) = \frac{q(x_t \mid x_{t-1})\, q(x_{t-1} \mid x_0)}{q(x_t \mid x_0)}$$

جایگذاری فرمول‌های گاوسی: حالا هر یک از این توزیع‌ها را با فرمول‌های گاوسی جایگزین می‌کنیم:

$$= \frac{\frac{1}{\sqrt{2\pi(1-\alpha_t)}} \exp\Big(-\frac{(x_t-\sqrt{\alpha_t}x_{t-1})^2}{2(1-\alpha_t)}\Big) \cdot \frac{1}{\sqrt{2\pi(1-\bar{\alpha}_{t-1})}} \exp\Big(-\frac{(x_{t-1}-\sqrt{\bar{\alpha}_{t-1}}x_0)^2}{2(1-\bar{\alpha}_{t-1})}\Big)}{\frac{1}{\sqrt{2\pi(1-\bar{\alpha}_t)}} \exp\Big(-\frac{(x_t-\sqrt{\bar{\alpha}_t}x_0)^2}{2(1-\bar{\alpha}_t)}\Big)}$$

توضیح هر بخش:

- صورت کسر اول: $q(x_t|x_{t-1}) = \mathcal{N}(x_t; \sqrt{\alpha_t}x_{t-1}, (1-\alpha_t)I)$
- صورت کسر دوم: $q(x_{t-1}|x_0) = \mathcal{N}(x_{t-1}; \sqrt{\bar{\alpha}_{t-1}}x_0, (1-\bar{\alpha}_{t-1})I)$
- مخرج: $q(x_t|x_0) = \mathcal{N}(x_t; \sqrt{\bar{\alpha}_t}x_0, (1-\bar{\alpha}_t)I)$

ساده‌سازی با حذف ثابت‌ها: از آنجا که ما فقط به شکل توزیع علاقه‌مند هستیم، می‌توانیم ثابت‌های نرمال‌سازی را نادیده بگیریم:

$$\propto \exp\Bigg(- \frac{(x_t - \sqrt{\alpha_t}x_{t-1})^2}{2(1-\alpha_t)} - \frac{(x_{t-1} - \sqrt{\bar{\alpha}_{t-1}}x_0)^2}{2(1-\bar{\alpha}_{t-1})} + \frac{(x_t - \sqrt{\bar{\alpha}_t}x_0)^2}{2(1-\bar{\alpha}_t)} \Bigg)$$

چرا جمله آخر مثبت است؟ چون در مخرج کسر قرار داشت، وقتی آن را به صورت لگاریتمی می‌نویسیم، علامت منفی اضافی می‌گیرد که کل علامت را مثبت می‌کند.

باز کردن مربعات: حالا باید مربعات را باز کنیم و عبارت را بر حسب x_{t-1} مرتب کنیم:

$$\begin{aligned} = \exp\Bigg(-\frac{1}{2}\Bigg[&\frac{1}{1-\alpha_t}\big(x_t^2 - 2\sqrt{\alpha_t}x_t x_{t-1} + \alpha_t x_{t-1}^2\big) \\ &+ \frac{1}{1-\bar{\alpha}_{t-1}}\big(x_{t-1}^2 - 2\sqrt{\bar{\alpha}_{t-1}}x_{t-1}x_0 + \bar{\alpha}_{t-1}x_0^2\big) \\ &- \frac{1}{1-\bar{\alpha}_t}\big(x_t^2 - 2\sqrt{\bar{\alpha}_t}x_t x_0 + \bar{\alpha}_t x_0^2\big)\Bigg]\Bigg) \end{aligned}$$

جمع‌آوری جملات بر حسب x_{t-1}: چون متغیر تصادفی این توزیع x_{t-1} میباشد پس برای اینکه این عبارت را به شکل استاندارد گاوسی در بیاوریم، باید جملات را بر حسب توان‌های x_{t-1} گروه‌بندی کنیم:

$$\propto \exp\Bigg(-\frac{1}{2}\Bigg[\left(\frac{\alpha_t}{1-\alpha_t} + \frac{1}{1-\bar{\alpha}_{t-1}}\right) x_{t-1}^2 \\ - 2\left(\frac{\sqrt{\alpha_t}}{1-\alpha_t}x_t + \frac{\sqrt{\bar{\alpha}_{t-1}}}{1-\bar{\alpha}_{t-1}}x_0\right) x_{t-1} + \text{ثابت‌ها}\Bigg]\Bigg)$$

تبدیل به شکل استاندارد گاوسی: این عبارت را می‌توان به شکل $\exp\left(-\frac{1}{2}A(x_{t-1} - \frac{B}{A})^2\right)$ نوشت که در آن:

$$A = \frac{\alpha_t}{1-\alpha_t} + \frac{1}{1-\bar{\alpha}_{t-1}}$$
$$B = \frac{\sqrt{\alpha_t}}{1-\alpha_t}x_t + \frac{\sqrt{\bar{\alpha}_{t-1}}}{1-\bar{\alpha}_{t-1}}x_0$$

محاسبه واریانس: واریانس توزیع گاوسی برابر است با $\frac{1}{A}$:

$$\mathrm{Var}[x_{t-1} \mid x_t, x_0] = \frac{1}{A} = \frac{1}{\frac{\alpha_t}{1-\alpha_t} + \frac{1}{1-\bar{\alpha}_{t-1}}}$$

ساده‌سازی واریانس: با انجام محاسبات جبری:

$$\begin{aligned}\frac{1}{A} &= \frac{1}{\frac{\alpha_t(1-\bar{\alpha}_{t-1})+(1-\alpha_t)}{(1-\alpha_t)(1-\bar{\alpha}_{t-1})}} \\ &= \frac{(1-\alpha_t)(1-\bar{\alpha}_{t-1})}{\alpha_t(1-\bar{\alpha}_{t-1})+(1-\alpha_t)} \\ &= \frac{(1-\alpha_t)(1-\bar{\alpha}_{t-1})}{\alpha_t-\alpha_t\bar{\alpha}_{t-1}+1-\alpha_t} \\ &= \frac{(1-\alpha_t)(1-\bar{\alpha}_{t-1})}{1-\alpha_t\bar{\alpha}_{t-1}}\end{aligned}$$

استفاده از رابطه $\bar{\alpha}_t = \alpha_t\bar{\alpha}_{t-1}$:

$$\begin{aligned}\mathrm{Var}[x_{t-1} \mid x_t, x_0] &= \frac{(1-\alpha_t)(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t} \\ &\equiv \tilde{\beta}_t\end{aligned}$$

محاسبه میانگین: میانگین توزیع گاوسی برابر است با $\frac{B}{A}$:

$$\begin{aligned}E[x_{t-1} \mid x_t, x_0] &= \frac{B}{A} \\ &= \left(\frac{\sqrt{\alpha_t}}{1-\alpha_t}x_t + \frac{\sqrt{\bar{\alpha}_{t-1}}}{1-\bar{\alpha}_{t-1}}x_0\right)\cdot\frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t}(1-\alpha_t) \\ &= \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}x_t + \frac{\sqrt{\bar{\alpha}_{t-1}}(1-\alpha_t)}{1-\bar{\alpha}_t}x_0 \\ &\equiv \tilde{\mu}_t\end{aligned}$$

نتیجه نهایی:

$$q(x_{t-1} \mid x_t, x_0) = \mathcal{N}\big(x_{t-1};\, \tilde{\mu}_t,\, \tilde{\beta}_t\big)$$

خلاصه فرمول‌های به دست آمده

پس همانطور که نشان دادیم داریم:

$$\tilde{\beta}_t = \frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t}\cdot(1-\alpha_t) = \frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t}\cdot(\beta_t)$$
$$\tilde{\boldsymbol{\mu}}_t(\mathbf{x}_t,\mathbf{x}_0) = \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}\mathbf{x}_t + \frac{\sqrt{\bar{\alpha}_{t-1}}(\beta_t)}{1-\bar{\alpha}_t}\mathbf{x}_0$$

بازنویسی میانگین بر حسب نویز

و چون داشتیم: $\mathbf{x}_t = \sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}_t$

حل برای x_0: پس می‌توان با جابجایی ترم‌ها نوشت: $\mathbf{x}_0 = \frac{1}{\sqrt{\bar{\alpha}_t}}(\mathbf{x}_t - \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}_t)$

توضیح این جابجایی: این رابطه نشان می‌دهد که اگر نویز موجود در x_t (یعنی ϵ_t) را بدانیم، می‌توانیم تصویر اصلی x_0 را بازیابی کنیم.

جایگذاری در فرمول میانگین: پس می‌توان میانگین $q(x_{t-1} \mid x_t, x_0)$ را که محاسبه کردیم را بازنویسی کنیم:

$$\tilde{\boldsymbol{\mu}}_t = \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}\mathbf{x}_t + \frac{\sqrt{\bar{\alpha}_{t-1}}(\beta_t)}{1-\bar{\alpha}_t}\frac{1}{\sqrt{\bar{\alpha}_t}}(\mathbf{x}_t - \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}_t)$$

ساده‌سازی جبری: با انجام محاسبات جبری و ساده‌سازی:

$$\tilde{\boldsymbol{\mu}}_t = \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}\mathbf{x}_t + \frac{\sqrt{\bar{\alpha}_{t-1}}(\beta_t)}{1-\bar{\alpha}_t}\cdot\frac{1}{\sqrt{\bar{\alpha}_t}}\mathbf{x}_t$$
$$- \frac{\sqrt{\bar{\alpha}_{t-1}}(1-\alpha_t)}{1-\bar{\alpha}_t}\cdot\frac{\sqrt{1-\bar{\alpha}_t}}{\sqrt{\bar{\alpha}_t}}\boldsymbol{\epsilon}_t$$

$$
\begin{aligned}
\text{ضریب}\, x_t &= \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t} + \frac{\sqrt{\bar{\alpha}_{t-1}}(1-\alpha_t)}{(1-\bar{\alpha}_t)\sqrt{\bar{\alpha}_t}} \\
&= \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1}) + \sqrt{\bar{\alpha}_{t-1}}(1-\alpha_t)/\sqrt{\bar{\alpha}_t}}{1-\bar{\alpha}_t} \\
&= \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1}) + \sqrt{\bar{\alpha}_{t-1}}(1-\alpha_t)/\sqrt{\bar{\alpha}_t}}{1-\bar{\alpha}_t}
\end{aligned}
$$

تجمیع ضرایب x_t:

با استفاده از $\bar{\alpha}_t = \alpha_t \bar{\alpha}_{t-1}$: پس از محاسبات طولانی که شامل استفاده از رابطه $\bar{\alpha}_t = \alpha_t \bar{\alpha}_{t-1}$ می‌شود، نتیجه نهایی به دست می‌آید:

$$\tilde{\boldsymbol{\mu}}_t = \frac{1}{\sqrt{\alpha_t}}\Big(\mathbf{x}_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\boldsymbol{\epsilon}_t\Big)$$

اهمیت این فرمول نهایی: این فرمول مهم است چون نشان می‌دهد که:

- برای دینویز کردن x_t، فقط باید نویز ϵ_t را تخمین بزنیم
- میانگین فرایند دینویزینگ به صورت خطی به تخمین نویز وابسته است
- این فرمول اساس طراحی شبکه عصبی دینویزر است که ϵ_t را پیش‌بینی می‌کند

فرایند تخمین زدن

همانطور که دیدیم نمیتوانیم که $q(x_{t-1}|x_t)$ را به صورت مستقیم محاسبه کنیم پس باید توسط مدل p_θ تخمین بزنیمش ما با فرض اینکه بتا ها به اندازه کافی کوچک بودند داشتیم که میتوان فرایند معکوس کردن دینویزینگ هم گاوسی در نظر گرفت پس در هر مرحله فرض میکنیم که مدل ما هم به صورت گاوسی فرایند معکوس کردن را تخمین میزند یعنی:

$$p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t) = \mathcal{N}(\mathbf{x}_{t-1}; \boldsymbol{\mu}_\theta(\mathbf{x}_t, t), \boldsymbol{\Sigma}_\theta(\mathbf{x}_t, t))$$

و برای اینکار نیز دو ورودی میگرد یک تصویر نویزی شده در زمان t و خود زمان t که با استفاده از این بتوانیم انعطاف پذیری مدل را در تخمین نویز ها را بیشتر بکنیم. همان فرض مستقل بودن ایجاد نویز هارا که در زمان t زده بودیم را برای مستقل بودن تخمین نویز هارا نیز لحاظ میکنیم پس بنابراین همانطوری که قبلا برای $q(x_t|x_{t-1})$ لحاظر کرده بودیم خواهیم داشت:

$$p_\theta(\mathbf{x}_{0:T}) = p(\mathbf{x}_T)\prod_{t=1}^{T} p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t)$$

که در اینجا $p(x_T)$ نویز خالص گاوسی هست.

خب حالا برای اپتیمایز کردن این فرایند چکار میشود انجام داد ؟ خب ما میخواهیم که با تابع تخمین زننده امون اگر عکس x_0 ای دادیم در اخر احتمال اینکه با اون عکس برسیم ماکسیموم بشود پس ایده اولی که به ذهن میرسد ماکسیمایز کردن log likelihood میتواند باشد یعنی اینکه ما این معادله را ماکسیمایز کنیم

$$p_\theta(x_0) = \int p_\theta(x_{0:T})d\mathbf{x_{1:T}}$$

اما این کار عملی نیست زیرا که ما باید تمام trajectories را از حالت نویز خالص گاوسی تا موقعی که به تصویر در حالت x_0 میرسیم محاسبه کنیم که همچین کاری را عملا نمیتوان انجام داد.

پس ایده بعدی چه میتواند باشد؟ اینکه از Variational Lower Bound استفاده کنیم

حالا باید بفهمیم که شهود این روش چیست؟

فرض کنید فرآیندی به صورت $z \to x$ داریم که می‌تواند نمونه‌های داده x را از متغیر z تولید کند. متغیر z می‌تواند شامل اطلاعاتی دربارهٔ ویژگی‌های یک تصویر باشد و از طریق این فرآیند، که با $p(x|z)$ نمایش میدهیم، این اطلاعات را میتوان به به تصویر واقعی x تبدیل کرد.

ما علاقه‌مندیم که فرآیند معکوس را بدانیم، یعنی این‌که چگونه از روی x، z را به دست آوریم. اگر توزیع z را بدانیم، می‌توانیم از قانون بیز استفاده کنیم:

$$p(z|x) = \frac{p(z)p(x|z)}{\int p(z,x)\,dz}.$$

اما به‌طور کلی نمی‌دانیم که چگونه انتگرال $\int p(z,x)\,dz$ را محاسبه کنیم. بنابراین نمی‌توانیم مخرج این عبارت را مستقیماً محاسبه کنیم و به جای آن توزیع $p(z|x)$ را با تابعی به شکل $q_\phi(z|x)$ تقریب می‌زنیم.

این تابع می‌تواند از یک از خانواده از توابعی که تابع مارا تخمین میزنند یعنی $\mathcal{Q}$ انتخاب شود، جایی که ما به دنبال انتخاب عضوی هستیم که یک معیار فاصله Δ را بین q_ϕ و توزیع واقعی p کمینه کند. داریم:

$$\hat{q}_\phi(z|x) = \arg\min_{q \in \mathcal{Q}} \Delta\big(q_\phi(z|x) \| p(z|x)\big). \qquad (۱۵)$$

برای مثال، اگر $\mathcal{Q}$ را مجموعهٔ تمام توزیع‌های گاوسی در نظر بگیریم، آنگاه $\phi = \{\mu, \Sigma\}$ پارامترهایی هستند که هر عضو این مجموعه را مشخص می‌کنند، و معادلهٔ (۱۵) به صورت انتزاعی توصیف می‌کند که چگونه باید بهترین پارامترها انتخاب شوند. واگرایی کولبک–لیبلر (D_{KL}) میتوان به عنوان یک معیار فاصلهٔ مطلوب در نظر گرفته شود، زیرا خواص ریاضی خوبی دارد.

شهود این واگرایی این هست که می آید و فاصله دو توزیع را محاسبه میکند برای اینکه چرا همچین چیزی را محاسبه میکند میتوانید از یکی از منابع نظریه اطلاعات این موضوع را مطالعه کنید.

تعریف این واگرایی این است:

$$D_{KL}(q \| p) = \int q(x) \, \log \frac{q(x)}{p(x)} \, dx = E_{x \sim q} \left[\log \frac{q(x)}{p(x)} \right]$$

حالا همین ایده را برای تابع تخمین زننده توزیع معکوس خودمون استفاده میکنیم و طبق روابط تعریف شده داریم:

$$\begin{aligned}
\log p(x_0) &= \log \int p(x_{0:T}) \, dx_{1:T} \\
&= \log \int \frac{p(x_{0:T}) \, q(x_{1:T} \mid x_0)}{q(x_{1:T} \mid x_0)} \, dx_{1:T} \\
&= \log E_{q(x_{1:T}|x_0)} \left[\frac{p(x_{0:T})}{q(x_{1:T} \mid x_0)} \right] \\
&\geq E_{q(x_{1:T}|x_0)} \left[\log \frac{p(x_{0:T})}{q(x_{1:T} \mid x_0)} \right] \qquad \text{Jensen inequality} \\
&= E_{q(x_{1:T}|x_0)} \left[\log \frac{p(x_T) \prod_{t=1}^{T} p_\theta(x_{t-1} \mid x_t)}{\prod_{t=1}^{T} q(x_t \mid x_{t-1})} \right]
\end{aligned}$$

خب ما نمیتوانستیم log likelihood را به طور مستقیم به دست بیاریم ولی حالا میتوانیم likelihood log را توسط این نابرابری ای که به دست اوردیم به سمت ماکسیمم اش حرکت بدهیم حالا باید طرف دیگر نابرابری را ماکسیم یا بهینه کنیم زیرا با این کار تابعی که میخواستیم ماکسیمایز کنیم یعنی احتمال اینکه تولید تصویر واقعی x_0 نیز توسط این نابرابری ماکسیمم میشود

دقت کنید که اینکار شدنی هست زیرا ما از توزیع هایی که فرم بسته اشان را نداشتیم problem را به توزیع هایی که فرم انالتیک یا بسته اشان را داریم تقلیل دادیم

همچنین اگر معادله اخر را بسط و ادامه دهیم داریم:

$$\begin{aligned}
\log p_\theta(x_0) \geq L_{VLB} = E_{q(x_{1:T}|x_0)}&\left[\log \frac{p(x_T)\prod_{t=1}^{T} p_\theta(x_{t-1} \mid x_t)}{\prod_{t=1}^{T} q(x_t \mid x_{t-1})}\right] \\
&= E_{q(x_{1:T}|x_0)}\big[\log p_\theta(x_0 \mid x_1)\big] + E_{q(x_{1:T}|x_0)}\left[\log \frac{p_\theta(x_T)}{q(x_T \mid x_{T-1})}\right] \\
&\quad + \sum_{t=1}^{T-1} E_{q(x_{1:T}|x_0)}\left[\log \frac{p_\theta(x_t \mid x_{t+1})}{q(x_t \mid x_{t-1})}\right] \\
&= E_{q(x_1|x_0)}\big[\log p_\theta(x_0 \mid x_1)\big] + E_{q(x_{T-1:T}|x_0)}\left[\log \frac{p_\theta(x_T)}{q(x_T \mid x_{T-1})}\right] \\
&\quad + \sum_{t=1}^{T-1} E_{q(x_{t-1:t+1}|x_0)}\left[\log \frac{p_\theta(x_t \mid x_{t+1})}{q(x_t \mid x_{t-1})}\right].
\end{aligned}$$

همچنین داریم:

$$E_{q(x_1|x_0)}\big[\log p_\theta(x_0 \mid x_1)\big] = \log p_\theta(x_0 \mid x_1)$$

یعنی lower bound ما برابر است با:

$$-E_q[\underbrace{D_{\mathrm{KL}}(q(\mathbf{x}_T|\mathbf{x}_0) \parallel p_\theta(\mathbf{x}_T))}_{L_T} - \sum_{t=2}^{T} \underbrace{D_{\mathrm{KL}}(q(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_0) \parallel p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t))}_{L_{t-1}} \underbrace{+ \log p_\theta(\mathbf{x}_0|\mathbf{x}_1)}_{L_0}]$$

به این خاطر منفی پشت ترم های ما میاید زیرا جایگاه انها در لاگ در معادله برعکس میشود.

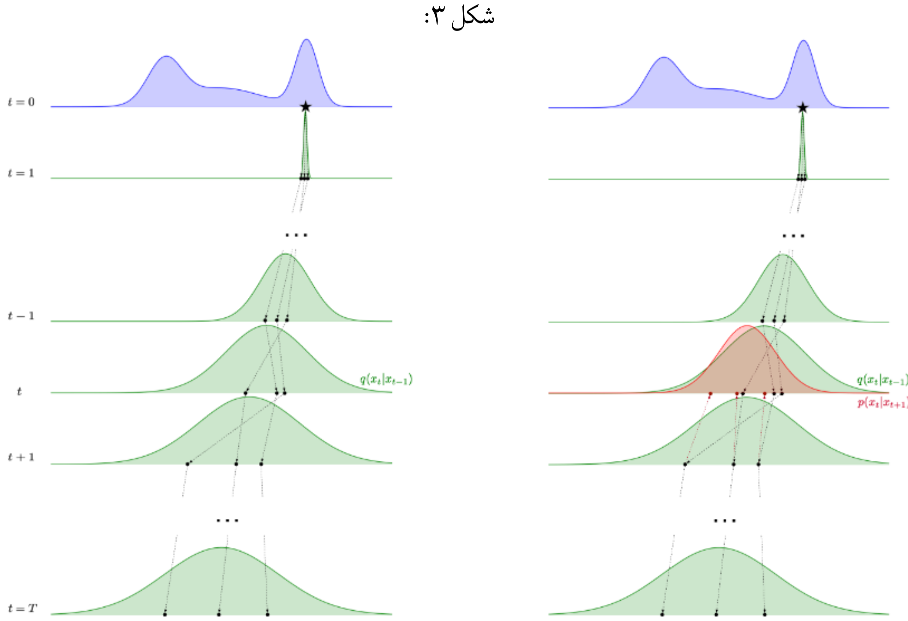
زیرا $log(x/y) = -log(y/x)$

حالا ماکسیمم کردن یک ترم با مینیموم کردن منفی ان معادل است پس ما میخواهیم این عبارت را مینیموم کنیم:

$$\begin{aligned}
L_{\text{VLB}} &= L_T + L_{T-1} + \cdots + L_0 \\
\text{where} L_T &= D_{\text{KL}}(q(\mathbf{x}_T|\mathbf{x}_0) \parallel p_\theta(\mathbf{x}_T)) \\
L_t &= D_{\text{KL}}(q(\mathbf{x}_t|\mathbf{x}_{t+1}, \mathbf{x}_0) \parallel p_\theta(\mathbf{x}_t|\mathbf{x}_{t+1})) \text{ for } 1 \leq t \leq T-1 \\
L_0 &= -\log p_\theta(\mathbf{x}_0|\mathbf{x}_1)
\end{aligned}$$

برای مینمایز کردن باید در نظر داشته باشیم که میتوانیم L_T را نادیده بگیریم زیرا هر دو عبارت صورت و مخرج کسر توزیع خالص گاوسی هستند پس بنابر این خیلی به هم نزدیک هستند و فاصله اشان برابر با صفر میشود و همچنین میتوان L_0 را نادیده گرفت زیرا تاثیر زیادی در نزدیک شدن تابع تخمین زننده ما به توزیع معکوس ندارد پس در واقع ما باید L_t هارا مینیموم کنیم اگر به یاد داشته باشید ما $p_\theta(x_{t-1}|x_t)$ را ارگیومنت اوردیم که در واقع یک توزیع گاوسی هست

شکل ۳:



پس حالا چکار میتوانیم بکنیم که

$$p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t) = \mathcal{N}(\mathbf{x}_{t-1}; \boldsymbol{\mu}_\theta(\mathbf{x}_t, t), \boldsymbol{\Sigma}_\theta(\mathbf{x}_t, t))$$

فاصله اش با

$$q(x_{t-1} \mid x_t, x_0) = \mathcal{N}\big(x_{t-1};\, \tilde{\mu}_t,\, \tilde{\beta}_t\big)$$

به حالت مینیموم یعنی L_t را مینیموم کنیم؟

چون که واریانس های $q(x_{t-1}|x_t, x_0)$ ثابت هستند واریانس تابع تخمین زننده را هم میتوانیم همین واریانس در نظر بگیریم.

ولی برای میانگین چطور؟

داشتیم که :

$$\tilde{\boldsymbol{\mu}}_t = \frac{1}{\sqrt{\alpha_t}}\Big(\mathbf{x}_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\boldsymbol{\epsilon}_t\Big)$$

پس میخواهیم که میانگین تابع تخمین زننده ما هم به میانگین توزیع معکوس نزدیک باشد یعنی میانگین را داشته باشیم:

$$\begin{aligned}
\boldsymbol{\mu}_\theta(\mathbf{x}_t, t) &= \frac{1}{\sqrt{\alpha_t}}\Big(\mathbf{x}_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\boldsymbol{\epsilon}_\theta(\mathbf{x}_t, t)\Big) \\
\mathbf{x}_{t-1} &= \mathcal{N}(\mathbf{x}_{t-1}; \frac{1}{\sqrt{\alpha_t}}\Big(\mathbf{x}_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\boldsymbol{\epsilon}_\theta(\mathbf{x}_t, t)\Big), \boldsymbol{\Sigma}_\theta(\mathbf{x}_t, t))
\end{aligned}$$

پس باید در واقع فاصله میانگین های تابع تخمین زننده را با میانگین توزیع های معکوس مینیموم کنیم پس:

$$L_t = E_{\mathbf{x}_0, \boldsymbol{\epsilon}}\Big[\frac{1}{2\|\boldsymbol{\Sigma}_\theta(\mathbf{x}_t, t)\|_2^2}\|\tilde{\boldsymbol{\mu}}_t(\mathbf{x}_t, \mathbf{x}_0) - \boldsymbol{\mu}_\theta(\mathbf{x}_t, t)\|^2\Big]$$

و تجربه نشان داده که ساده سازی تابع بالا به تابع پایین عملکرد ترینینگ را بهبود میبخشد پس با بهینه سازی تابع پایین ما مدل را ترین میکنیم و اینگونه میتوان فرایند بازگشتی را از نویز به تصویر ترین کرد.

$$\begin{aligned} L_t^{\text{simple}} &= E_{t\sim[1,T],\mathbf{x}_0,\boldsymbol{\epsilon}_t}\left[\|\boldsymbol{\epsilon}_t - \boldsymbol{\epsilon}_\theta(\mathbf{x}_t, t)\|^2\right] \\ &= E_{t\sim[1,T],\mathbf{x}_0,\boldsymbol{\epsilon}_t}\left[\|\boldsymbol{\epsilon}_t - \boldsymbol{\epsilon}_\theta(\sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}_t, t)\|^2\right] \end{aligned}$$

در واقع شهود تابع بالا این هست که ما میخواهیم دینویزری را به دست بیاوریم که به طور متوسط در هر مرحله تخمین خوبی از کل فرایند دینوزینگ و تشخیص نویز های وارد شده را به ما بدهد.

کاربرد عملی در آموزش مدل

نحوه آموزش:

۱. یک تصویر x_0 از دیتاست انتخاب می‌کنیم
۲. یک زمان تصادفی t انتخاب می‌کنیم
۳. نویز گاوسی $\epsilon \sim \mathcal{N}(0, I)$ تولید می‌کنیم
۴. تصویر نویزی $x_t = \sqrt{\bar{\alpha}_t}x_0 + \sqrt{1-\bar{\alpha}_t}\epsilon$ محاسبه می‌کنیم
۵. شبکه عصبی سعی می‌کند ϵ را از روی x_t و t تخمین بزند
۶. تابع هزینه: $L = ||\epsilon - \epsilon_\theta(x_t, t)||^2$

نحوه نمونه‌برداری (Sampling):

۱. از نویز کامل $x_T \sim \mathcal{N}(0, I)$ شروع می‌کنیم
۲. برای $t = T, T-1, ..., 1$:

- نویز را با شبکه آموزش‌دیده تخمین می‌زنیم: $\hat{\epsilon} = \epsilon_\theta(x_t, t)$
- میانگین را محاسبه می‌کنیم: $\mu = \frac{1}{\sqrt{\alpha_t}}(x_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\hat{\epsilon})$
- $x_{t-1} \sim \mathcal{N}(\mu, \tilde{\beta}_t I)$ نمونه‌برداری می‌کنیم

۳. نتیجه نهایی x_0 تصویر تولید شده است

در ادامه اگر وقت کردم این نوت نوشته شده را تبدیل به یک پست وبلاگ interactive میکنم و همچنین

مدل های score based را شرح میدهم و همچنین guided Diffusion را هم توضیح میدم

منابع:

Denoising Diffusion Probabilistic Models Jonathan Ho et el

Generative Modeling by Estimating Gradients of the Data Distribution Yang Song, Stefano Ermon

Lecture Notes in Probabilistic Diffusion Models , Inga Strumke and Helge Langseth

Weng, Lilian. (Jul 2021). What are diffusion models? Lil'Log. https://lilianweng.github.io/posts/2021-07-11-diffusion-models/.

Note on Diffusion Generative Models , https://animadversio.github.io/ academic$_n otes/note - on - diffusion - generative - model$