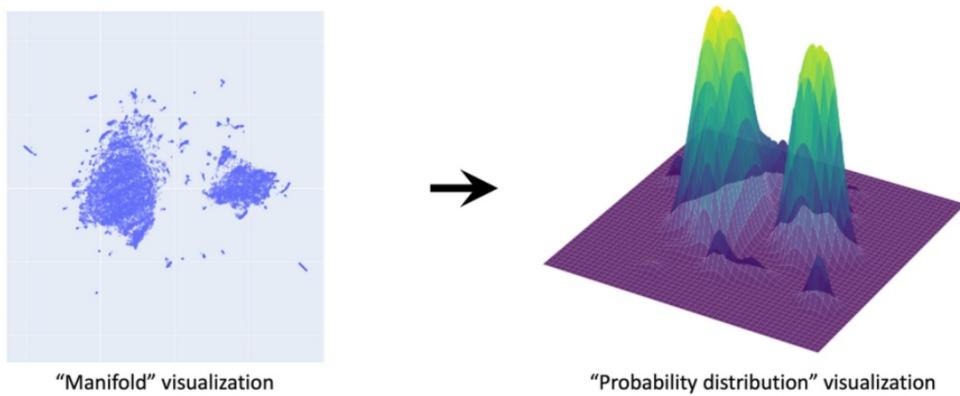


Всё, что делает генеративный AI — изучает распределение, а затем синтезирует из него

В случае набора изображений, мы считаем, что отдельно взятый образец x извлечён из некоторого распределения в пиксельном пространстве.

Если у нас есть изображения $128 \times 128 (= 16384 \text{ пикс})$, то их реальное распределение $q(x)$ лежит в пространстве размерностью 16384. Но благодаря гипотезе многообразия мы можем понизить размерность и её структуру распределения сохранится. Гипотеза предполагает, что существенные наборы данных лежат на многообразиях более низкой размерности, вложенных в пространство более высокой размерности (плоскость, вложенная в 3D-пространство). Есть много способов понижения размерности. Они дают понять, как примерно выглядит распред. $q(x)$.

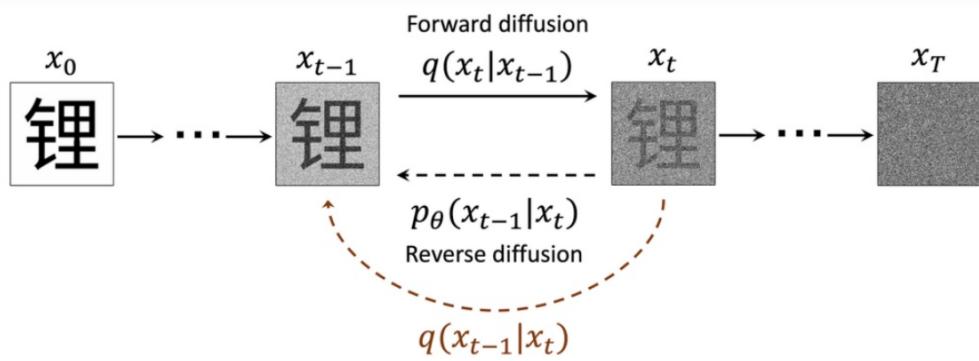


Для более простых распределений мы можем просто вычислить параметры (пр.: для одномерного гауссово-вереско и дисперсию). Диффузионные модели в свою очередь не параметризуют распред. напрямую, а учатся ему косвенно, через обучение предсказываемого шума в данных.

Что такое шум?

Шум в диффуз. моделях — гауссовой шум. Каждый пиксель в "чисто зашумленном" изображении — число из распред. $\mathcal{N}(0, 1)$.
На практике это многомерное распред. $\mathcal{N}(0, I)$, I -единичная матрица (матрица ковариации), означающая отсутствие корреляции между шумом в разных пикселях. Искоррелированный шум.

Обзор диффузионного процесса



Прямая диффузия

Этот процесс соответствует преобразованию, которое приближает изобр. к чистому шуму путем уменьшения значений пикселей и добавления гауссового шума, центрированного на значениях пикселей прошлого изображения. Записывается как $q(x_t|x_{t-1})$, где "t" указывает, что средние значения из $q(x_t)$ передаются перед

Маркированное распределение $q(x_t)$ накапливает эффект всех пред. шагов зашумления.

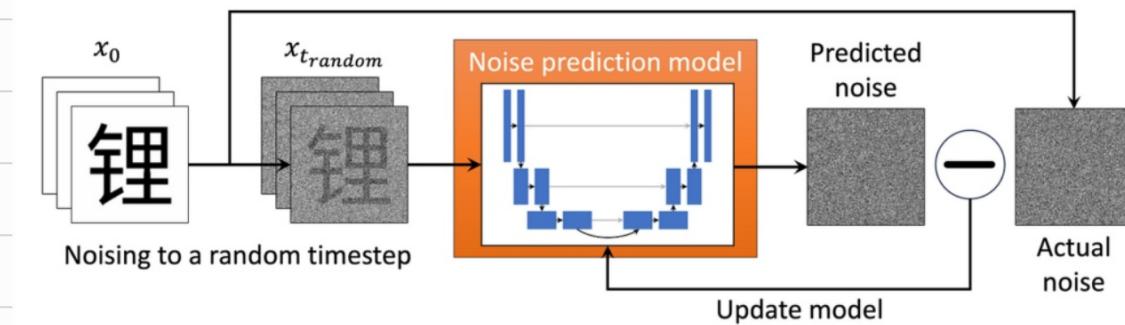
Каждому временному шагу соответствует дисперсия для шума распред. (напр. 0,0001 — 0,02 за 1000 шагов, variance scheduler)

Зашумление не нужно выполнять последовательно — для любого исходного образца $x_0^{(i)}$ можем вычислить зашумленный результат $x_T^{(i)}$, просто симплифицировав $q(x_t|x_0=x_0^{(i)})$.

Обратная диффузия

Если бы мы знали обратное распред. $q(x_{t-1}|x_t)$, то, начав с образца чистого шума x_t , мы могли повторно вычитать порции шума, пока не получим $x_0 \sim q(x_0)$. На практике, не зная x_0 , мы не можем получить доступ к этому распред., поэтому прибегаем к ML-модели для предсказания шума

Для предсказания шума используется U-Net.



В ходе обучения модель нейроно изучает обратное диффузионное распред.. Это позволяет модели симулировать из распредел. $q(x_0)$ постепенным удалением шума

В процессе обучения используется большое кол-во шагов ($T=1000$), чтобы траектория шум \rightarrow образец была очень плавной. При генерации используется меньшее кол-во шагов.

Есть более продвинутые методы симуляции и в индустрии часто симулирование переформулируется через дифференциальные уравнения