Diffusion原理

一、模型流程

训练过程

模型的训练主要包含两部分，一个是噪声添加部分，一个是去噪部分：

从模型的原始思路上看噪声添加部分具体流程如下：

噪声添加部分（原始）：

1. 随机选取一张图x以及时间步t
2. 从高斯分布中采样噪声，添加到图片上，得到新的图片
3. 不断重复第2步t次得到一张图片与最后一次的噪声ε

但这样一步一步添加噪声显然太慢了，通过概率推导，我们可以将多次添加的高斯噪声融合到一起，具体来说我们第t次添加的噪声占比为，即

我们可令

则公式会变成这样

这样的话我们就可以得到新的模型训练过程了：

噪声添加部分（实际）：

1. 随机选取一张图x以及时间步t
2. 从高斯分布中采样噪声ε，通过公式得到新的图片以及对应ε

再配上去噪部分，我们的模型训练流程即可完成

去噪部分：

1. 得到t与加上噪声的图片
2. 讲二者（,t）输入U Net模型，得到第t步添加的噪声
3. 对比和ε间的差距，反向传播训练模型

采样流程

模型的使用流程如下：

1. 采样得到高斯分布
2. 对上一步得到的图片进行去噪工作，对高斯分布采样得到（增加随机性用，t=1时让采样结果为0），使用模型预测出噪声，
3. 重复上述步骤t次，直至得到，这就是我们生成的图片

二、数学原理

这一部分主要聚焦于我们生成图片时，为什么要用下面的奇怪公式：

生成模型的目标

我们其实可以这么理解图片生成模型:我们训练的模型其实是一个转换器，这个转换器的目标是将高斯分布（采样所得）转化为图片的真实分布(转化所得)。

那么我们要如何量化自身的目标呢？有两种出发点，一种是最大似然，另一种是计算KL散度，但最后的结果是一样的，我们都要最大化同一个数值：

首先，我们用模型转化出的分布被记为，其中是真实图片，我们的目标是

即使得模型产生真实图片的概率最大。

其实这个最大化也可以等同与

即

接下来让我们转向DDPM的整体过程

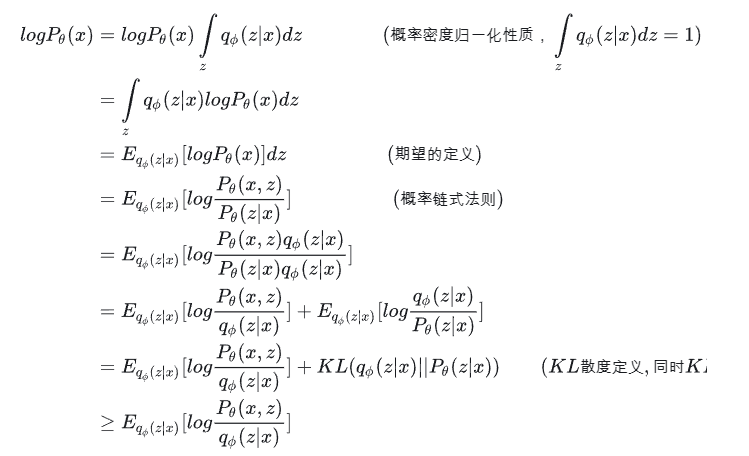
DDPM的生成过程

首先让我们梳理下模型训练与使用过程中的各个操作表达为公式具体应该是什么样子：

我们的diffusion模型在使用的过程中的实质是将转化为（表现为预测噪声并用噪声处理加噪图片为去噪图片），即计算：，在将作为原始输入的情况下，这个分布可以写作

而在训练过程中，我们加噪过程也是一个转换：将转换为，这样的分布记为q，即q(),如果是在的基础上得到的，我们的分布可以写作q()

接下来让我们推导



这一部分经过图中这一系列复杂的转换最终可以转化成

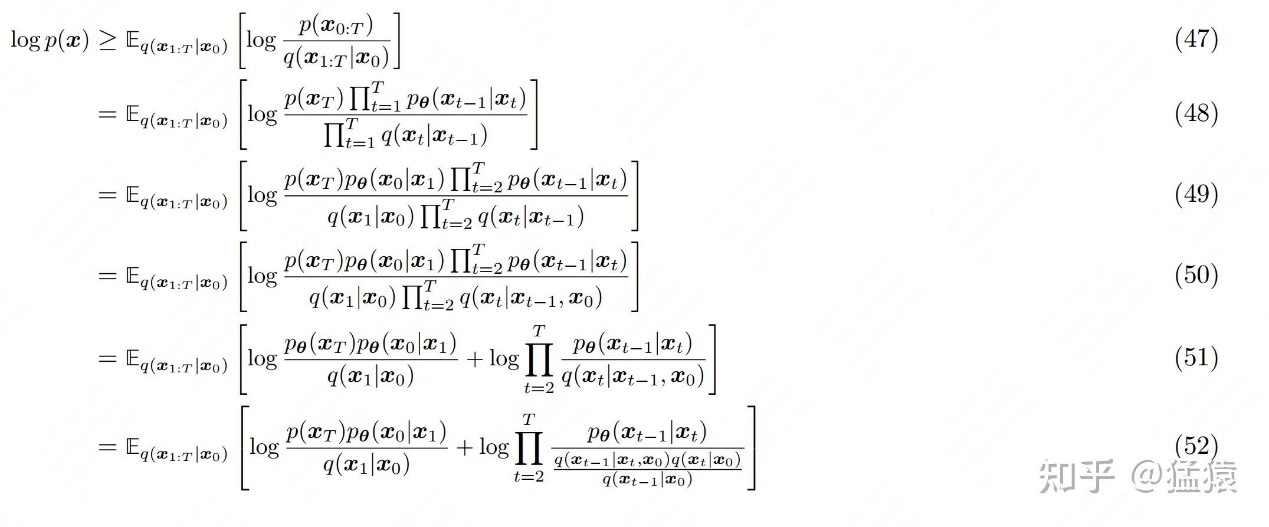
其中是永远大于0的，即

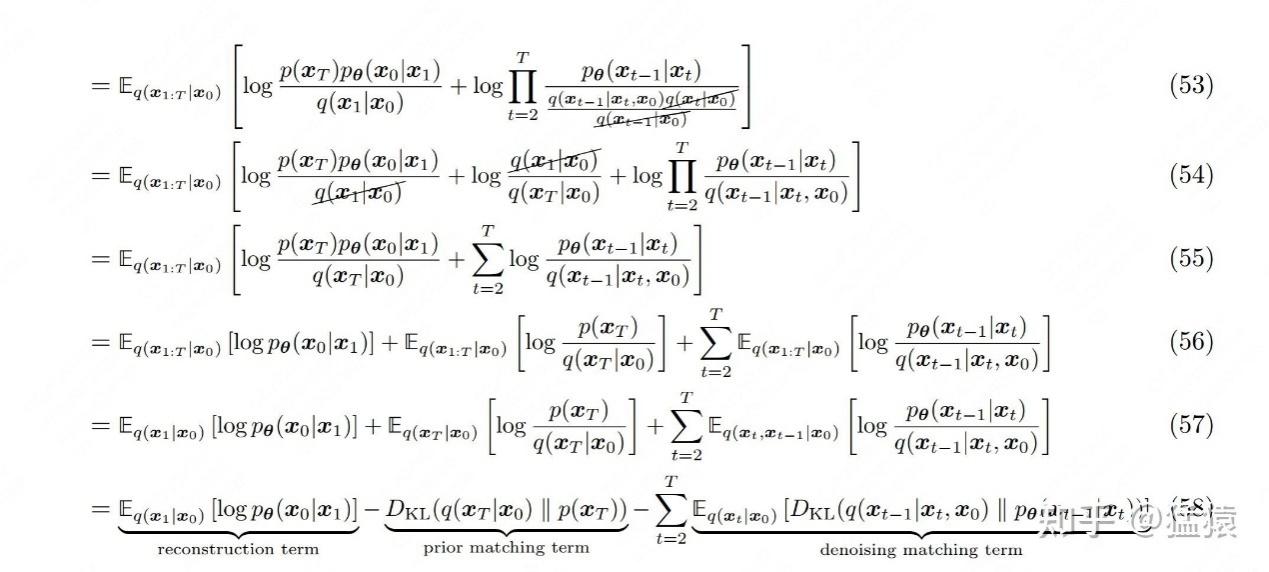
这样，我们实质上要最大化，即

在上面的推导中，我们的x代表真实图片，z代表此前的加噪图片，如果聚焦于某一张图片（训练实质上也是一个一个算的，又不是最优化求整体最优），使用

来表示他，我们可以这么表示上面的式子：

接下来我们继续推到上面的式子：





通过上图的推导，我们可以将优化目标转化为下面的内容：

其中第二项和θ无关，只有第一项和第三项是我们要关注的，让我们聚焦于第三项上，继续进行推导：

接下来，只要我们能分析出具体是什么情况，就能知道该如何修改分布了

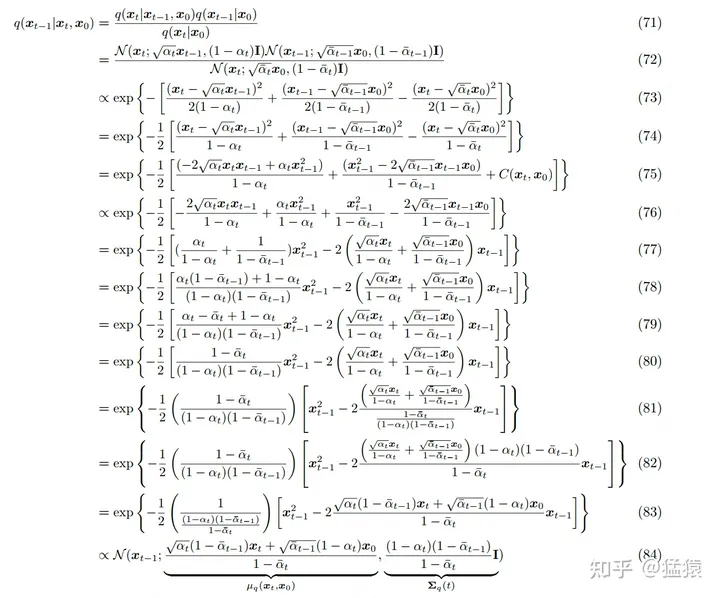
首先，让我们回顾下之前β与α等参数的内容，是第t次加噪时加噪前图片和加噪后图片间的信噪比衡量参数，随着t的增大，β也在不断增大，为了简化计算，我们将多次加噪合并为了对原图片进行一次加噪的操作，并β的公式用下面的变量进行表示：

并因此得到的计算方法：

因此我们可以知道中各个q中变量的关系

同时，由于我们的加噪过程就是逐渐将图片模糊为高斯分布，我们也就能知道三个q服从高斯分布，因此

高斯分布之间是可以进行合并的，因此，我们可以推导出下面的内容



因此，将用表示后。在这个分布中，它的方差与均值如下：

再让我们看一眼，对于分布而言，方差是已经被固定好的内容，我们要修改的核心其实是均值μ，也就是说，我们需要让我们模型转换得到的分布的的μ能接近，而在中，只有是变量，因此，只要我们的模型能预测噪声，我们就能让模型转化出来的分布接近于真实图片的分布

以上就是ddpm的生成原理，总的来说就是通过对噪声的拟合对给定噪声图片去噪从而生成符合数据分布的图片，而我们之前提到的那些更加复杂的模型也是基于这样的一个生成原理，只是我们会有条件作为引导来进一步调整去噪过程，具体如下：

**传统DDPM中的隐含条件输入：**

在传统的DDPM中，虽然模型没有显式的条件输入，但训练数据集本身确实提供了隐含的条件信息。训练数据集定义了模型学习的目标分布，模型通过学习训练数据集的统计特性来生成相似的图像。

可以认为，训练数据集中的图像风格、结构和内容是模型隐式学习的条件。模型通过这些隐含条件，在去噪过程中生成与训练数据集分布相似的图像。

对于需要结合图像和文字描述进行图像生成或转换的任务，例如将给定图像按照文字描述进行转变，扩散模型可以通过条件生成的方式来实现。这种技术通常涉及到在扩散模型的去噪过程或生成过程中的每一步中引入条件信息（图像和文字描述）。

# 分类器基本思想和作用

## 一、基本思想

在Diffusion Models Beat GANs on Image Synthesis发布之前，diffusion模型和GAN模型有一定的差距，作者认为这一方面是因为GAN模型开发的比较久，模型结构优化的好，另一方面是GAN模型有个判别器，能让模型能比较好的在多样性和真实性之间做区分，因此作者受到启发，给diffusion模型的生成过程也加了个分类器，希望能用分类器指导图片的生成：

使用不同时间步的图像训练一个分类器(记为，ε也这么干)，再使用分类器的梯度信息指导扩散模型的生成。

即在生成过程中，模型在计算完毕当前的均值与方差后，会用分类器计算梯度，并依据梯度修改得到新的均值.

## 数学推导

首先让我们确定一些结论（其中代表增加分类器指导后的概率分布，q为增加前的）：

1. 是确定值
2. = (加噪过程与类别无关,下面的4,5,6由此式推出，其实是说增加类别指导后，我们产生噪声的过程是和之前一样，不受影响的)
3. 分类器的y只与当前的状态有关，而与此前状态无关（论文已证），即
4. 在确定y与前一步含噪图片的情况下，产出图片的概率分布：

在增加了分类器指导后，我们的后验概率p(|)也出现了变化（是常数，与无关，因此令Z=，也就是说：

又因为:

对分布取log,带入，对进行泰勒展开，展开到一阶就行（据说t足够高的时候，接近于0，而的曲率<，因此不用管二阶导及以后的部分），即(这里的μ是估计出来的，g指的是二阶导)

则整体公式结果如下：

因此，依据如上推导，我们可以依据分类器的结果导数辅助指导我们的生成过程，具体流程如下：

1. 接受输入y与分类器指导权重s
2. 从N(0,I)中随机采样
3. 依据diffusion模型预测μ与∑
4. 依据分类的结果μ，∑，，从下面的分布中采样:
5. 重复3与4直到得到

## 好处

对于无条件的扩散模型，分类器梯度让模型能按类输出。

同时提升了模型质量：在足够高的比例下，引导的无条件模型可以非常接近未引导的条件模型的FID，引导条件模型进一步提高了FID。

可以通过调整分类器梯度的比例，让模型在多样性和真实性之间进行权衡。

分类器条件去噪有以下三个主要的好处，有效解决了无条件生成模型的缺陷：

1. 增强生成控制：通过引入分类器的梯度信息，生成过程得到了明确的目标导向，使生成结果更加符合预期的类别或特性。

2. 提高生成质量：分类器的引导信号帮助生成模型避免生成不相关或错误的特征，从而提高生成图像的质量和一致性。

3. 增加多样性：分类器指导方法利用分类器学习到的数据分布特征，能够在生成过程中保持和增强图像的多样性。