山西大学学报(自然科学版)48(1):120—129,2025
Journal of Shanxi University(Nat. Sci. Ed.)

DOI:10.13451/j.sxu.ns.2024140

基于选择状态空间的去噪扩散概率模型研究

佘志用[1],康家荣[2*],张东坡[1],郭晓新[3]
(1. 新疆政法学院 信息网络安全学院,新疆 图木舒克 844000;
2. 广西民族师范学院 数理与电子信息工程学院,广西 崇左 532200;
3. 吉林大学 计算机科学与技术学院,吉林 长春 130012)

摘 要:针对去噪扩散概率模型(DDPM)的采样效率低、训练时间长和硬件资源开销大等问题,提出了选择状态空间的去噪扩散概率模型。该方法首先用选择状态空间模型(SSM)的动态选择性来提高DDPM在长序列上的采样效率;其次通过多方向扫描图像,使得DDPM扩散时获取更多有效的图像信息;最后利用SSM的线性时间复杂以及并行运算减少DDPM训练时的时间和硬件资源的开销。DDPM、改进的扩散模型(DDIM)、变分自编码器的去噪扩散概率模型(VAE-DDPM)、视觉Transformer的去噪扩散概率模型(VIT-DDPM)和本文方法在ImageNet和人脸图像数据集(FFHQ)数据集进行图像生成实验时,对不同分辨率图像的Frechet距离(FID)、结构相似性(SSIM)、峰值信噪比(PSNR)和生成时间等参数对比分析,本文方法生成128×128图像时FID、SSIM和PSNR的值分别提升了5.6%~22.6%、4.7%~15.5%和1.9%~6.6%,得出该方法能够有效解决DDPM的缺陷,并且优于其他扩散模型。

关键词:扩散模型;线性时间复杂度;多方向扫描
中图分类号:TP391.4 **文献标志码:**A **文章编号:**0253-2395(2025)01-0120-10

Research on the Probability Model of Denoising Diffusion Based on Selection State Space

SHE Zhiyong[1], KANG Jiarong[2*], ZHANG Dongpo[1], GUO Xiaoxin[3]
(1. *College of Information Network Security, Xinjiang University of Political Science and Law, Tumushuk* 844000*, China;*
2. *College of Mathematics and Electronic Information Engineering, Guangxi Minzu Normal University, Chongzuo* 532200*, China;*
3. *College of Computer Science and Technology, Jilin University, Changchun* 130012*, China*)

Abstract: Aiming at the problems of low sampling efficiency, long training time, and high hardware resource overhead of the Denoising Diffusion Probability Model (DDPM), a state space based DDPM is proposed. First, this method uses the dynamic selectivity of the State Space Model (SSM) to improve the sampling efficiency of DDPM on long sequences; Second, by scanning the image in multiple directions, more effective image information can be obtained during the diffusion of DDPM; Finally, the time and hardware resource overhead during DDPM training were reduced by utilizing the linear time complexity and parallel computing of SSM. When conducting image generation experiments on ImageNet and Flickr-Faces-High-Quality (FFHQ) datasets using DDPM, Denoising Diffusion Implicit Models (DDIM), Variational Auto-Encoder-DDPM (VAE-DDPM), Vision Transformer-DDPM (VIT-DDPM), and our proposed method, a comparative analysis was conducted on parameters such as Frechet Inception Distance (FID), Structural Similarity Index (SSIM), Peak Signal-to-Noise Ratio (PSNR), and generation time of images with different resolutions. When generating 128×128 images, the values of FID, SSIM, and PSNR of our method increased by 5.6%-22.6%, 4.7%-15.5%, and 1.9%-6.6%, re-

收稿日期:2024-06-17;**接受日期:**2024-10-24
基金项目:国家自然科学基金(82071995);新疆政法学院校长基金(XZZK2021002;XZZK2022008)
作者简介:佘志用(1990—),男,甘肃张掖人,硕士,讲师,研究方向为图像处理。E-mail:szysoft378@126.com
* **通信作者:**康家荣(KANG Jiarong),E-mail:821464008@qq.com
引文格式:佘志用,康家荣,张东坡,等.基于选择状态空间的去噪扩散概率模型研究[J].山西大学学报(自然科学版),2025,48(1):120-129. DOI:10.13451/j.sxu.ns.2024140.

spectively. It is concluded that this method can effectively solve the defects of DDPM and is superior to other diffusion models.
Key words: diffusion model; linear time complexity; multidirectional scanning

0 引言

目前主流的生成模型有对抗神经网络(Generative Adversarial Networks, GAN)[1]和扩散模型,因GAN在训练生成器和判别器时收敛困难已逐步被扩散模型取代。Ho等人提出的去噪扩散概率模型(Denoising Diffusion Probability Model, DDPM)[2]成为扩散模型领域的主流模型之一。DDPM对比GAN不但具有更详细的生成信息,而且生成的图像具有高质量的多样性。在非条件图像生成领域DDPM已经击败GAN,例如Dhariwal等在图像视觉领域提出扩散模型在图像合成中击败GAN[3]。DDPM在马尔可夫链长序列上采样导致采样效率比GAN要低,针对DDPM该缺陷很多学者在该方面对DDPM做了大量改进工作,在不损失生成图像质量的前提下提高采样效率。例如Nichol等提出非马尔可夫链的去噪扩散概率模型(Denoising Diffusion Implicit Models, DDIM)[4],选取采样子列提高采样效率,因采样子列是随机选取导致了生成图像的质量不稳定。冯月萍等提出粗糙集的去噪扩散概率模型[5]解决该问题。

DDPM图像生成的前期研究主要在图像空间上完成。当DDPM对高分辨图像在图像空间训练时,导致训练时间和硬件资源的开销变大。针对DDPM在图像空间扩散时的缺陷,有学者利用变分编码器[6](Variational Auto-Encoder, VAE)将DDPM从图像空间映射到隐空间进行图像扩散生成,提出了稳定扩散模型(Stable Diffusion)。例如Robin Rombach等提出基于隐空间扩散模型的高分辨率图像合成方法[7],该方法在隐空间上扩散生成高分辨率图像,有效减少模型训练时间和硬件资源开销,但该方法在隐空间上图像生成存在图像信息丢失的问题。

DDPM利用U-net网络训练时,通过局部感受野扫描图像获取图像信息,这样容易导致信息损失。针对该问题有学者利用注意力机制解决U-net网络局部感受野的问题,例如Chen等提出将空间和通道双注意与变压器U-net结合对医学图像分割的方法[8],William等提出可扩展的变压器扩散模型[9]。这些方法均在U-net的局部感受野加入注意力机制或Transformer获取图像重要信息,但DDPM中加入Transformer后模型时间效率上变为二次时间复杂度,导致DDPM推理效率变低。以上方法对DDPM改进和生成优势,使得DDPM在目标检测[10-11]、医学图像分割[12]、图像视频生成[13]和多模态等领域[14]得到广泛应用。

虽然对DDPM进行针对性的改进,但是并没有解决其本质问题。Gu等提出了具有选择性状态空间的线性时间序列建模方法(Mamba)[15],该方法在时间序列建模上具有训练快和推理快、长序列上具有良好的记忆功能、并行运算且时间效率为线性时间复杂度等优点,因此在序列建模上Mamba对比Rnn和Transformer具有很大优势,受VIT(Vision Transformer)方法的启发,Zhu等提出视觉状态空间模型方法(VMamba)[16-20],使得VMamba在视觉领域得到有效应用。根据Mamba优点以及在视觉领域的应用,本文提出了基于多方向扫描的Mamba-DDPM方法。该方法利用Mamba的优点解决DDPM中的缺陷,为了扩展感受野本文从多个方向扫描图像扩展到全局感受野且时间效率为线性时间复杂度,获取更多有效图像信息。Mamba和DDPM的有效结合使得扩散模型在序列建模、机器视觉、多模态和遥感图像处理等领域具有可靠性的应用前景。

1 相关工作

1.1 DDPM

DDPM包括前扩散和逆扩散过程如图1所示。

扩散过程中给定初始数据分布$X_0 \sim q(X_0)$,从$t=0$时刻到$t=T$时刻不断向对应的t添加高斯噪声,该噪声的标准差是以固定值β_t($\{\beta_t \in (0,1)\}_{t=1}^{T}$)来确定,均值是以固定值$\beta_t$和当前$t$时刻的数据$X_t$决定。且当$T \to \infty$则$\beta_1 < \beta_2 \cdots < \beta_t < \cdots < \beta_{T-1} < \beta_T$,整个扩散模型是一

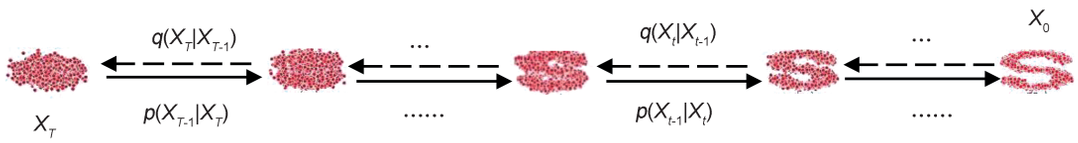


图1 扩散过程

Fig. 1 Diffusion process

个马尔可夫链过程。

DDPM某一时刻前扩散的条件概率分布如下：

$$q(X_t|X_{t-1})=N(X_t;\sqrt{1-\beta_t}\,X_{t-1},\beta_t I)。\quad(1)$$

通过X_0和α_t来表示任意时刻的$q(X_t)$，而不需要迭代，推导得出的结果如下所示：

$$X_t=\sqrt{\overline{\alpha}_t}\,X_0+\sqrt{1-\overline{\alpha}_t}\,Z。\quad(2)$$

根据参数重整化得出$q(X_t|X_0)=N(X_t;\sqrt{\overline{\alpha}_t}\,X_0,(1-\overline{\alpha}_t)I)$，其中$I$是标准正态分布的方差。

假设逆扩散过程也是马尔可夫链过程

$$p_\theta(X_{t-1}|X_t)=N\left(X_{t-1};\mu_\theta(X_t,t),\sum\nolimits_\theta(X_t,t)\right),$$

$$p_\theta(X_{0:T})=p(X_T)\prod_{t=1}^{T}p_\theta(X_{t-1}|X_t)。$$

由贝叶斯公式、式(1)、式(2)和概率密度函数推导损失函数(Loss)以及X_{t-1}的表达式：

$$\mathrm{Loss}=\|\varepsilon-\varepsilon_\theta(\sqrt{\overline{\alpha}_t}\,X_0+\sqrt{1-\overline{\alpha}_t}\,\varepsilon,t)\|^2,\quad(3)$$

$$X_{t-1}=\frac{1}{\sqrt{\alpha_t}}\left(X_t-\frac{\beta_t}{\sqrt{1-\overline{\alpha}_t}}\varepsilon_\theta(X_t,t)\right)+\sigma_t Z。\quad(4)$$

1.2 Mamba

选择状态空间模型来源于连续的状态方程如下所示：

$$\begin{aligned}h_t'&=Ah_{t-1}+Bx_{t-1},\\ y_t&=Ch_t+Dx_{t-1}。\end{aligned}\quad(5)$$

该方程中h_{t-1}、x_{t-1}、h_t'和y_t分别为前一时刻的状态、输入以及当前时刻的状态和输出。A、B、C和D分别为前一时刻的状态矩阵、输入矩阵以及当前时刻状态矩阵和输入矩阵。连续的状态方程的流程图如图2所示。

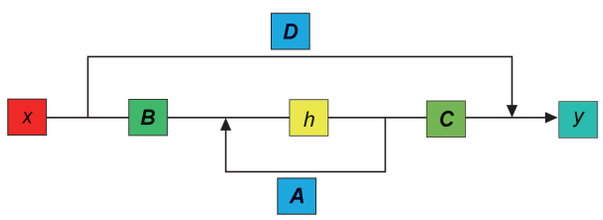


图2 连续的状态方程的流程图

Fig. 2 Flowchart of continuous state equations

公式(5)是处理连续的信号状态，对于自然语言、图像和视频等离散数据该公式无法处理。因此对公式(5)进行离散化得到离散的状态方程如下所示：

$$h_t=\overline{A}\,h_{t-1}+\overline{B}\,x_t,\ y_t=\overline{C}h_t,\quad(6)$$

$$\overline{A}=\exp(\Delta A),$$

$$\overline{B}=(\Delta A)^{-1}(\exp(\Delta A)-I)\cdot\Delta B。\quad(7)$$

离散状态方程的流程图如图3所示。

图3中的$\overline{A}$、$\overline{B}$、C、H、x和y分别表示变化状态矩阵、变化输入矩阵、前时刻状态矩阵、当前状态、输入内容和输出内容。公式(6)是时不变的状态方程，如果用该状态方程处理离散数据不具有选择性。为了让公式(6)具有选择性，从时不变改变为时变，这样得到时变状态

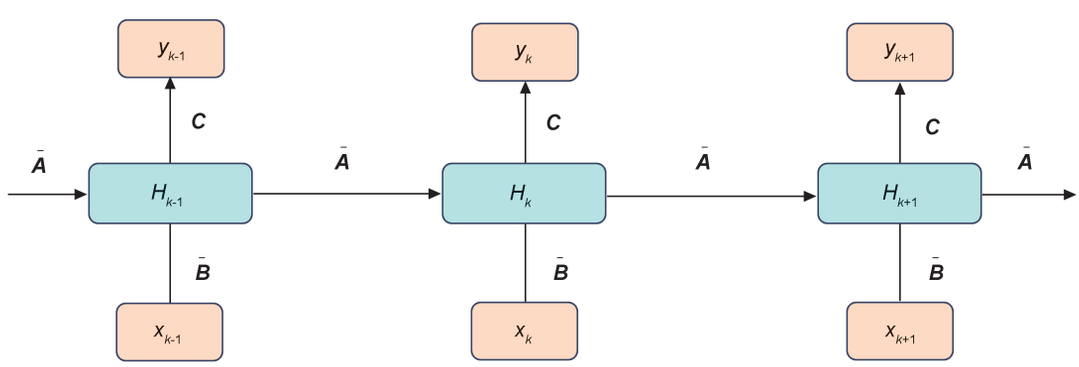


图3 离散状态方程的转态传递流程图

Fig. 3 Flow chart of transition of discrete state squations

方程如下所示:

$$h_t=s_{\overline{A}}(x_t)h_{t-1}+s_{\overline{B}}(x_t)x_t,$$
$$y_t=s_C(x_t)h_t。 \tag{8}$$

公式(8)就是Mamba的核心部分,该方程的不同矩阵变换如图4所示。

图4中左侧矩阵是输入内容矩阵x(L表示行数,D表示列,B表示矩阵个数),中间矩阵是变化状态矩阵$\overline{A}$(L表示行数,N表示列数,B表示矩阵个数),右侧是输入内容矩阵x与变化输入矩阵$\overline{B}$相乘所得矩阵。公式(8)是时变的离散方程不同时刻矩阵都在发生变化,因此无法利用卷积形式求解y。根据公式(6)和公式(8)迭代发现可以利用并行的方式求解y,流程图如图5所示。

图5是利用迭代并行运算求解状态H,图中首先计算出H_0和H_2,利用这两个状态分别计算出H_1和H_3。

2 方法描述

2D多方向扫描图像。为了解决DDPM局部感受野的问题,采用2D多方向扫描图像。受到VIT方法启发将输入的图像分割为均匀小块,分别对分割的图像进行行和列的扫描。行从左上和右下扫描,列从左下和右上扫描得到四个方向扫描结果如图6所示。VIT将2D维度的分割图像转化为1D的向量,图像包含了二维空间信息(如局部纹理和全局结构),如果将图像转为1D图像会失去二维空间信息。本文直接使用多方向扫描的2D图像输入不但扩展感受野,而且保留的图像二维空间信息。

用Mamba框架替换DDPM中的U-net或Transformer训练该模型。在DDPM从T到0时刻逆向扩散中从全高斯噪声图像逐步恢复图像过程中以公式(3)为目标函数,在Mamba框架上线性时间复杂度动态更新$\boldsymbol{A}$、$\boldsymbol{B}$和$\boldsymbol{C}$矩阵回复某一时刻的图像或是去噪。在整个序列上去噪恢复图像过程为了加快Mamba的推理效率,采用并行方法提高效率且改进GPU的运行结构提高硬件运行效率,如图7和图8所示。

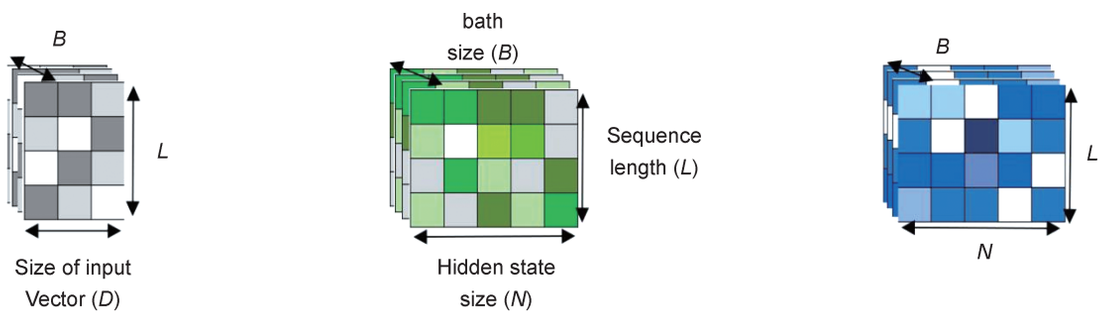


图4 公式(8)不同矩阵变换图

Fig. 4 Different matrix transformation graphs of formula (8)

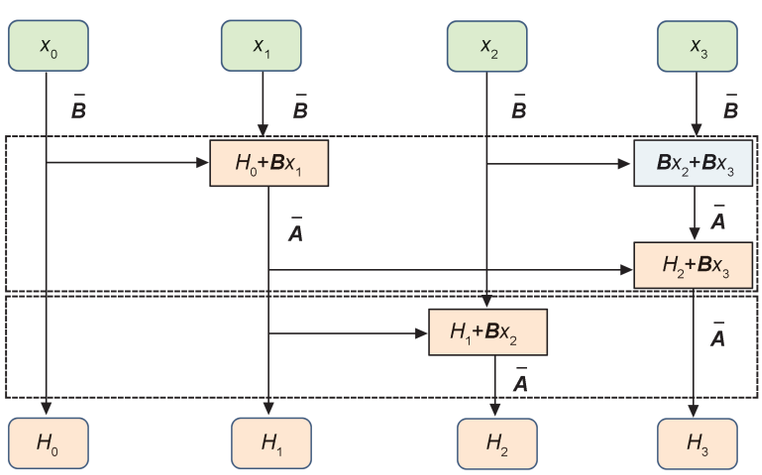


图5 公式(8)并行运算流程图

Fig. 5 Parallel operation flowchart of formula (8)

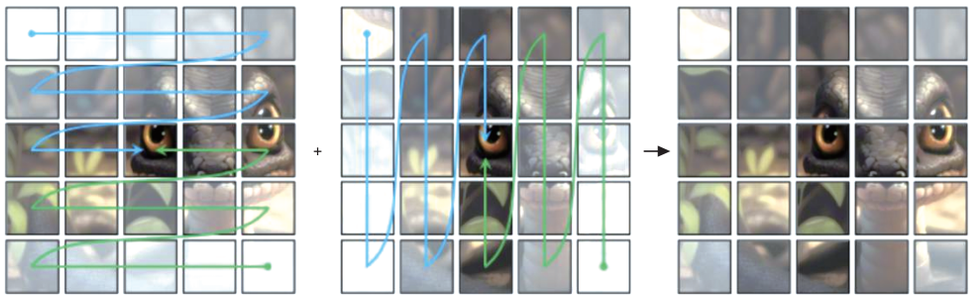
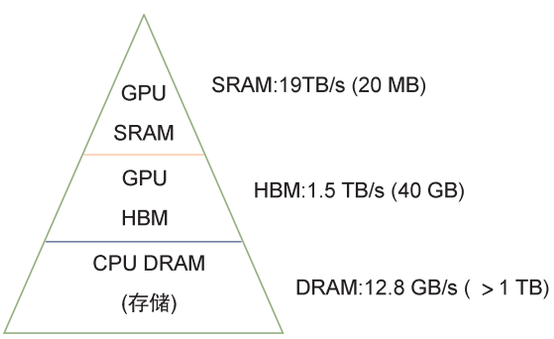


图6 2D多方向扫描图像

Fig. 6 Image of 2D multi-directional scanning

多方向扫描选择状态空间的去噪扩散概率模型(DDPM-Mamba)的框架图如图9所示。



注:SRAM(Static Random Access Memory,静态随机存储器);HBM(High Bandwidth Memory,高带宽内存);DRAM(Dynamic Random Access Memory,动态随机存取存储器)

图7 CPU和GPU运行与存储

Fig. 7 CPU and GPU operation and storage

图9是DDPM-Mamba方法的核心内容,因图像整体语义结构比较弱,无法在序列上整体处理图像,图9将图像分割为等大的子图。将分割的图像四个方向扫描得到序列,利用四个DDPM-Mamb模块在扫描序列上扩散生成结果,最后合并为完整的图像。

算法描述

DDPM-Mamba模型训练

DDPM-Mamba模型训练见算法1,各步骤解释如下:

Step 1:获取原图像X_0概率分;

Step 2:初始化时间t的取值;

Step 3:规定高斯噪声ε的取值;

Step 4:深度学习网络训练

Step 4.1:对X_t归一化处理得出B(bath)、L(sequence length)、D(vector);

Step 4.2:对$X_t^{'}$进行线性变换处理得到初始化的当前输入矩阵B_t;

Step 4.3:对$X_t^{'}$进行线性变换处理得到初始化当前时刻的状态C_t;

Step 4.4:得到变化量Δ_t;

Step 4.5:前一时刻变化的转态矩阵$\overline{A_{t-1}}$:(B,L,D,N);

Step 4.6:当前变化的输入矩阵$\overline{B_t}$:(B,L,D,N);

Step 4.7:$X_t^{'}$通过选择状态模型(State Space Model, SSM)得到变换的Y_t:(B,L,D);

Step 4.8:利用$\text{Loss}=\|\varepsilon-\varepsilon_\theta(\sqrt{\overline{\alpha}_t}X_0+\sqrt{1-\overline{\alpha}_t}\varepsilon,t)\|^2$循环在噪声误差范围内得到最优的$Y_t$:$(B,L,D)$;

Step 5:训练完毕。

DDPM-Mamba采样:

DDPM-Mamba采样见算法2,各步骤解释如下:

Step 1:$X_T\sim N(0,I)$属于标准正态分布;

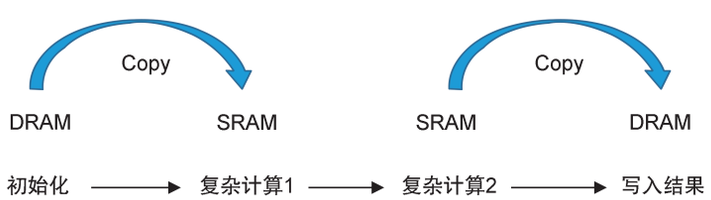


图8 改进GPU运行结构

Fig. 8 Improving operating structure of GPU

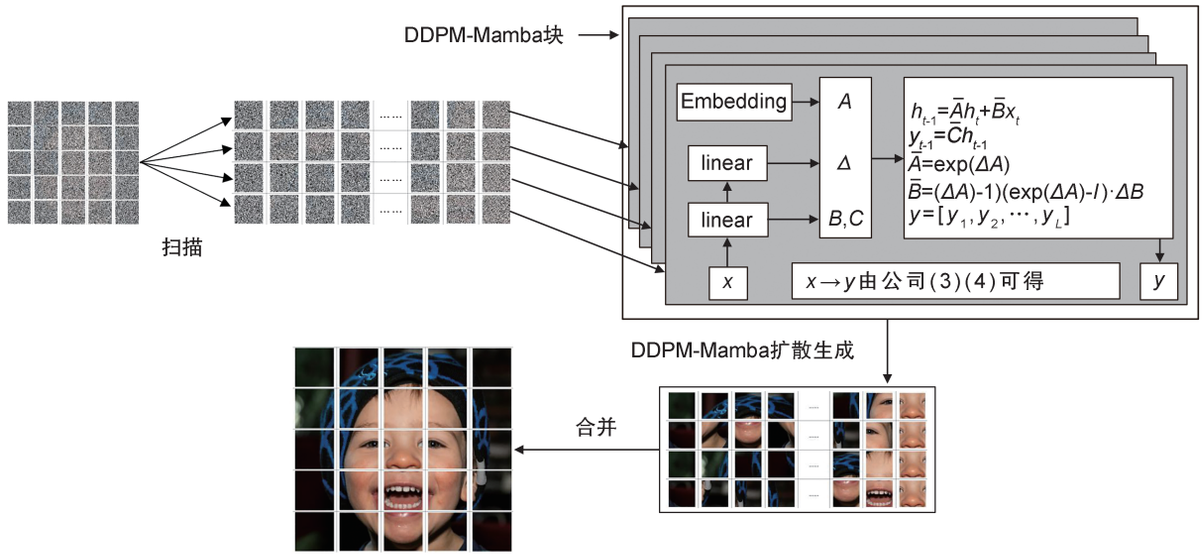


图9 DDPM-Mamba框架

Fig. 9 Framework of DDPM-Mamba

算法1 用公式(3)训练DDPM-Mamba网络

Step 1:X_0的概率分布$q(X_0)$

Step 2:$t\in\{1,2,\cdots,T-1,T\}$

Step 3:ε是从标准正态分布$N(0,I)$中随机获取。

Step 4:深度学习网络训练。

Step 4.1:输入序列归一化Norm$(X_t)\to X_t^{'}:(B,L,D)$

Step 4.2:Linear(LU(Convld$_t(X_t^{'})))\to B_t:(B,L,N)$

Step 4.3:Linear(LU(Convld$_t(X_t^{'})))\to C_t:(B,L,N)$

Step 4.4:log$(1+\exp(\text{linear}\Delta_t(X_t^{'})+\text{Parameter}\Delta_t))\to\Delta_t$:$(B,L,D)$

Step 4.5:$\Delta_t\otimes\text{Parameter}A_{t-1}\to\overline{A_{t-1}}:(B,L,D,N)$其中Parameter$A_{t-1}(D,N)$

Step 4.6:$\Delta_t\otimes B_t\to\overline{B_t}:(B,L,D,N)$

Step 4.7:SSM$(\overline{A_{t-1}},\overline{B_t},C_t)(X_t^{'})\to Y_t:(B,L,D)$

Step 4.8: Loss$=\|\varepsilon-\varepsilon_\theta(\sqrt{\bar{\alpha}_t}X_0+\sqrt{1-\bar{\alpha}_t}\varepsilon,t)\|^2$

Step 5:训练完毕

算法2 DDPM-Mamba采样算法

Step 1: $X_T\sim N(0,I)$

Step 2: Linear$(X_t^{'})\to z:(B,L,D)$

Step 3: for $t\in(T,0)$ in {左上,右下,左下,右上}do

Step 3.1:将$X_t^{'}$输入训练好的DDPM-Mamba模型中

Step 3.2:$Y_{t左上}\cdot\text{LU}(z)\to Y^{'}_{t左上}:(B,L,D)$

Step 3.3: $Y_{t右下}\cdot\text{LU}(z)\to Y^{'}_{t右下}:(B,L,D)$

Step 3.4:$Y_{t左下}\cdot\text{LU}(z)\to Y^{'}_{t左下}:(B,L,D)$

Step 3.5:$Y_{t右上}\cdot\text{LU}(z)\to Y^{'}_{t右上}:(B,L,D)$

Step 3.6:Linear$(Y^{'}_{t左上}+Y^{'}_{t右下}+Y^{'}_{t左下}+Y^{'}_{t右上})\to Y_t$

Step 3.7: $Y_{t-1}=\dfrac{1}{\sqrt{\alpha_t}}\left(Y_t-\dfrac{\beta_t}{\sqrt{1-\bar{\alpha}_t}}\varepsilon_\theta(Y_t,t)\right)+\sigma_tZ$

Step 4: end for

Step 5:return Y_0

Step 2:Linear$(X_t^{'})\to z:(B,L,D)$线性变换得到B(bath)、L(sequence length)、D(vector);

Step 3:在$t\in(T,0)$上循环多方向扫描得到{左上,右下,左下,右上}的Y_t

Step 3.1:将$X_t^{'}$输入训练好的DDPM-Mamba模型中;

Step 3.2:通过$Y_{t左上}\cdot\text{LU}(z)$得到$Y^{'}_{t左上}:(B,L,D)$;

Step 3.3:通过$Y_{t右下}\cdot\text{LU}(z)$得到$Y^{'}_{t右下}:(B,L,D)$;

Step 3.4:通过$Y_{t左下}\cdot\text{LU}(z)$得到$Y^{'}_{t左下}:(B,L,D)$;

Step 3.5:通过$Y_{t右上}\cdot\text{LU}(z)$得到$Y^{'}_{t右上}:(B,L,D)$;

Step 3.6:对Linear$(Y^{'}_{t左上}+Y^{'}_{t右下}+Y^{'}_{t左下}+Y^{'}_{t右上})$进行线性变换得到$Y_t$;

Step 3.7:利用

$$Y_{t-1}=\frac{1}{\sqrt{\alpha_t}}\left(Y_t-\frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}}\varepsilon_\theta(Y_t,t)\right)+\sigma_tZ$$

得到Y_{t-1};

Step 4:结束循环。

Step 5:返回利用DDPM-Mamba模型得到恢复的原图Y_0。

3 实验分析

本文实验在ImageNet和人脸图像数据集

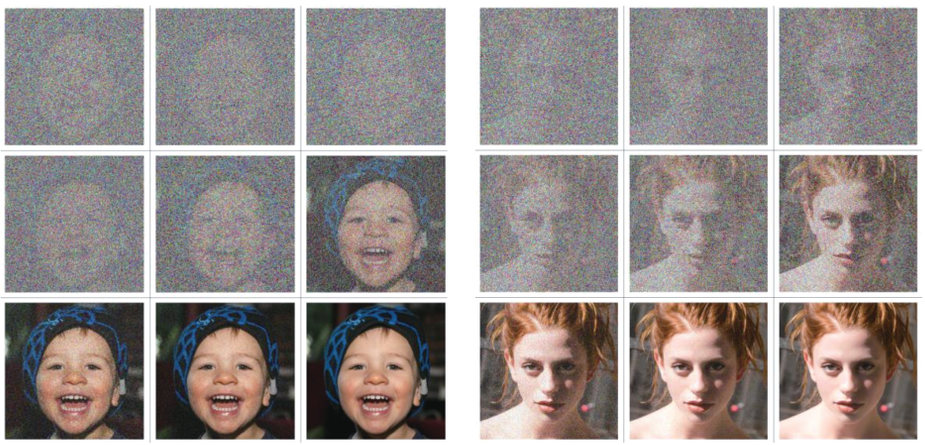


图10 DDPM-Mamba部分训练过程图

Fig. 10 Partial training process of DDPM-Mamba

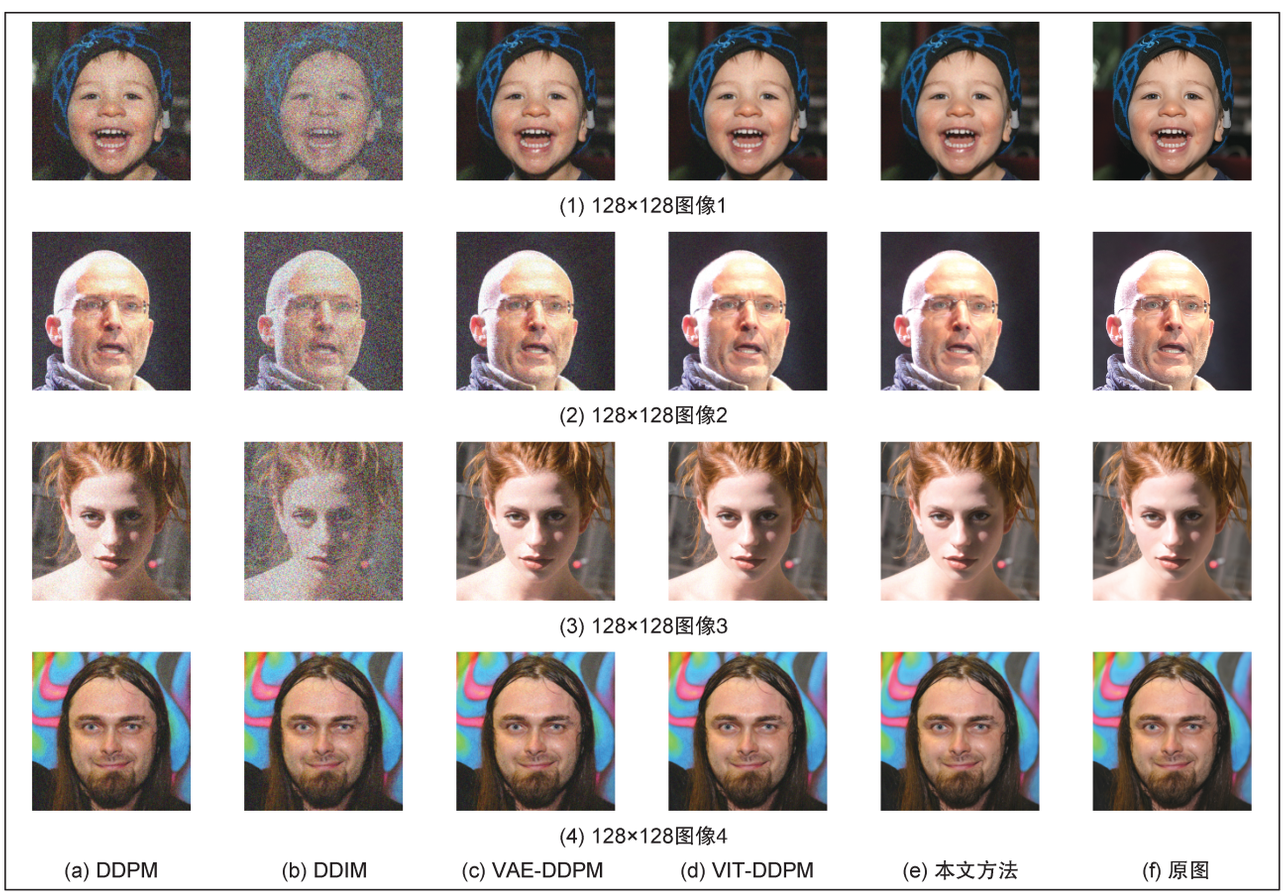


图11 五种方法生成128×128图像的效果图

Fig. 11 Five methods for generating renderings of 128×128 images

(Flickr-Faces-High-Quality,FFHQ)上进行对比分析,用到pytorch(GPU版本)深度学习框架,RTX A5000的显卡。根据选取的数据集和超参数 $\alpha_t(\{\alpha_t\in(0,1)\}_{t=1}^{T})$ 确定 $T=1\ 000$。DDPM-Mamba训练的过程图像如图10所示。

DDPM、DDIM、变分自编码器的去噪扩散概率模型(Variational Auto-Encoder-DDPM,VAE-DDPM)、视觉Transformer的去噪扩散概率模型(Vision Transformer-DDPM,VIT-DDPM)和本文DDPM-Mamba分别生成128×128和512×512的效果图如图11和图12所示。

通过Frechet距离(Frechet Inception Distance,FID)、结构相似性(Structural Similarity Index,SSIM)和峰值信噪比(Peak Signal-to-Noise

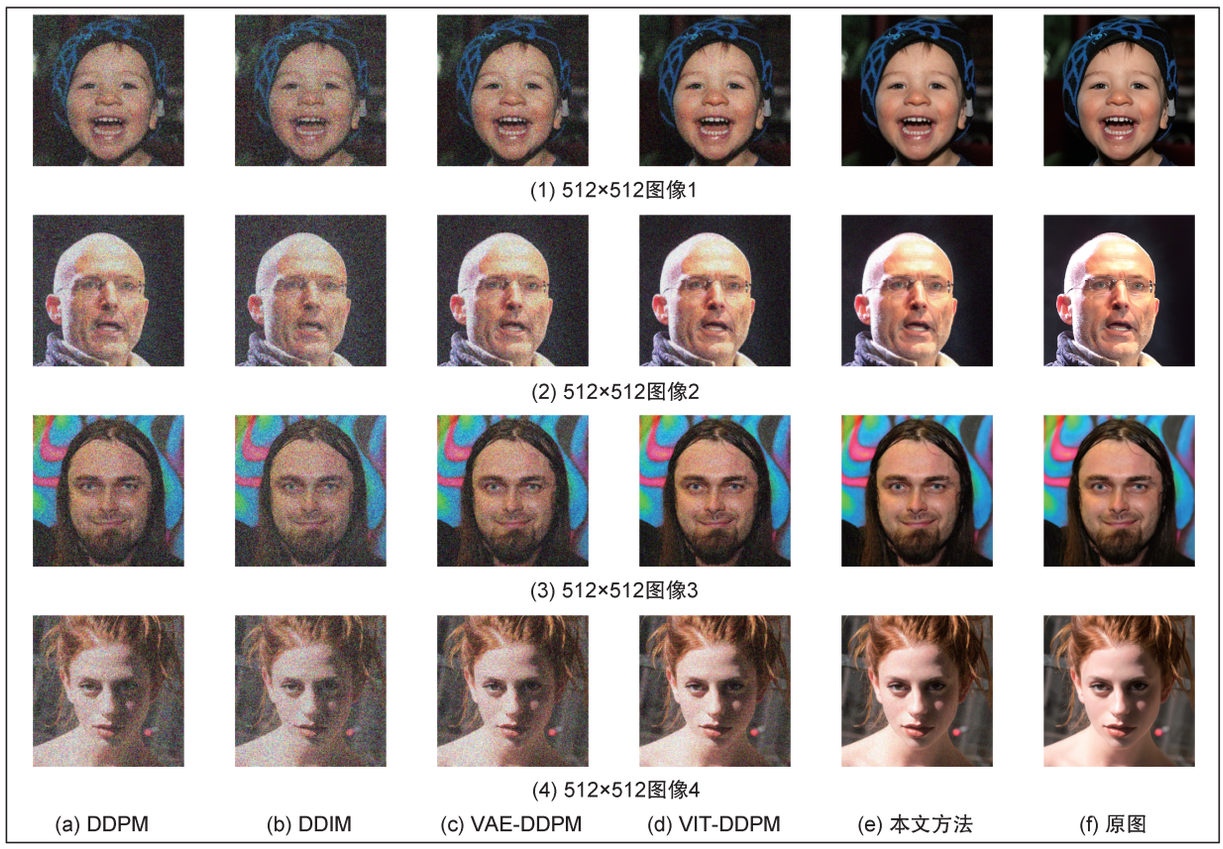


图12 五种方法生成512×512图像的效果图

Fig. 12 Five methods for generating renderings of 512×512 images

表1 图11效果图的FID、SSIM和PSNR的值

Table 1 The values of FID, SSIM, and PSNR for the rendering of Fig.11

序号	图像评估指标	(a) DDPM	(b) DDIM	(c) VAE-DDPM	(d) VIT-DDPM	(e) 本文方法
(1)	FID	30.42	43.33	28.13	22.23	**18.12**
	SSIM	0.74	0.71	0.77	0.77	**0.87**
	PSNR	28.02	27.57	28.86	28.88	**29.45**
(2)	FID	15.29	23.01	10.87	10.62	**10.05**
	SSIM	0.66	0.64	0.71	0.71	**0.84**
	PSNR	28.75	26.31	29.62	29.7	**31.8**
(3)	FID	20.84	32.39	14.10	14.02	**12.23**
	SSIM	0.76	0.75	0.81	0.81	**0.85**
	PSNR	28.90	26.44	29.74	29.83	**30.03**
(4)	FID	33.67	44.84	26.43	26.11	**23.07**
	SSIM	0.78	0.72	0.82	0.82	**0.83**
	PSNR	28.36	27.43	29.14	29.54	**29.88**

注:表中黑体数字表示区分本文方法与对比方法生成图像的指标参数。

表2 图12效果图的FID、SSIM和PSNR的值

Table 2 The values of FID, SSIM, and PSNR for the rendering of Fig.12

序号	图像评估指标	(a) DDPM	(b) DDIM	(c) VAE-DDPM	(d) VIT-DDPM	(e) 本文方法
(1)	FID	207.23	319.99	115.79	75.02	**21.98**
	SSIM	0.47	0.33	0.72	0.86	**0.87**
	PSNR	13.06	10.50	24.43	26.12	**26.80**
(2)	FID	221.69	338.51	107.25	70.12	**22.52**
	SSIM	0.49	0.36	0.60	0.75	**0.79**
	PSNR	10.83	9.66	12.85	22.56	**22.64**
(3)	FID	219.91	309.16	90.67	55.25	**21.02**
	SSIM	0.60	0.37	0.77	0.81	**0.86**
	PSNR	18.54	12.07	26.70	27.44	**27.63**
(4)	FID	223.45	312.56	95.86	63.20	**19.58**
	SSIM	0.52	0.41	0.68	0.79	**0.82**
	PSNR	14.23	10.89	22.40	26.85	**27.06**

注:表中黑体数字表示区分本文方法与对比方法生成图像的指标参数。

Ratio,PSNR)三个值量化评价五种方法生成不同分辨率图像的质量,图11和12效果图对应的量化表如表1和表2所示。

五种方法在相同的采样序列长度上(1 000)生成一张图像时间开支对比,如表3所示。

表3 采样时间对比

Table 3 Comparison of sampling time

方法	128分辨率图像耗时/s	512分辨率图像耗时/s
DDPM	53	151
DDIM	41	124
VAE-DDPM	56	182
VIT-DDPM	62	175
本文方法	**29**	**46**

表1和2中的FID值越小说明图像质量越好,而SSIM和PSNR值越大说明生成图像的质量越好。在五种方法生成128图像时通过表1分析本文方法FID相对较小,而SSIM和PSNR相对较大说明在该分辨率上本文方法生成的图像质量较好;在生成512图像时通过表2分析其他方法的FID普遍比较大,而本文方法较小;SSIM和PSNR其他方法均比较小,而本文方法比较大。说明在生成512分辨图像时其他四种方法均处于劣势,而本文方法能够生成质量较高图像。通过表3的生成不同分辨率图像时间对比,得到本文方法在相同序列长度下采样效率优于其他方法。最终得出本文提出的DDPM-Mamba模型在不同分辨率下生成图像不但图像质量优,而且采样效率高。

4 结论

本文通过去噪概率扩散模型、选择状态空间模型以及多方向扫描有效结合,在理论解决了DDPM采样效率低、局部感受野以及资源开支大等问题。DDPM-Mamba模型在分辨率为128和512上生成图像与DDPM、DDIM、VAE-DDPM和VIT-DDPM[21]对比得到该方法生成的图像质量均优于其他四种方法且采样效率较高。该方法在生成超分辨上具有很大潜力,但是该方法是否具备有效的条件扩散和生成图像更细特征还有待进一步研究。

参考文献:

[1] SUN Z, SHEN Y, ZHOU Q, *et al*. Principle-driven Self-alignment of Language Models from Scratch with Minimal Human Supervision[EB/OL]. (2023-05-04) [2024-06-08]. https://doi.org/10.48550/arXiv.2305.03047.

[2] HO J, JAIN A, ABBEEL P. Denoising Diffusion Probabilistic Models[EB/OL]. (2020-09-16) [2024-06-08]. https://arxiv.org/pdf/2006.11239.

[3] DHARIWAL P, NICHOL A. Diffusion Models Beat Gans on Image Synthesis[EB/OL]. (2021-06-01)[2024-06-08]. https://doi.org/10.48550/arXiv.2105.05233.

[4] NICHOL A, DHARIWAL P. Improved Denoising Diffusion Probabilistic Models[EB/OL]. (2021-02-18)[2024-06-08]. https://doi.org/10.48550/arXiv.2102.09672.

[5] 佘志用,郭晓新,冯月萍,等.基于粗糙集的去噪扩散概率方法[J].吉林大学学报(理学版),2024,**62**(2):339-346. DOI:10.13413/j.cnki.jdxblxb.2023139.
SHE ZHIYONG, GUO XIAOXIN, FENG YUEPING, *et al*.Research on the Probability Method of Denoising Diffusion Using Rough Sets[J]. *J Jilin Univ Sci Ed*,2024,**62**(2):339-346. DOI:10.13413/j.cnki.jdxblxb.2023139.

[6] GUU K, HASHIMOTO T B, OREN Y, *et al*. Generating Sentences by Editing Prototypes[J]. *Trans Assoc Comput Linguist*, 2018, **6**: 437-450. DOI: 10.1162/tacl_a_00030.

[7] ROMBACH R, BLATTMANN A, LORENZ D, *et al*. High-resolution Image Synthesis with Latent Diffusion Models[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New York: IEEE, 2022: 10674-10685. DOI: 10.1109/CVPR52688.2022.01042.

[8] CHEN B Z, LIU Y S, ZHANG Z, *et al*. TransAttUnet: Multi-level Attention-guided U-net with Transformer for Medical Image Segmentation[J]. *IEEE Trans Emerg Top Comput Intell*, 2024, **8**(1): 55-68. DOI: 10.1109/TETCI.2023.3309626.

[9] PEEBLES W, XIE S. Scalable Diffusion Models with Transformers[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. New York: IEEE. 2023:4195-4205. DO1:10.1109/ICCV51070.2023.00387.

[10] PINAYA W H L, GRAHAM M S, GRAY R, *et al*. Fast Unsupervised Brain Anomaly Detection and Segmentation with Diffusion Models[M]//Lecture Notes in Computer Science. Cham: Springer Nature Switzerland, 2022: 705-714. DOI: 10.1007/978-3-031-16452-1_67.

[11] CHEN S F, SUN P Z, SONG Y B, *et al*. DiffusionDet: Diffusion Model for Object Detection[C]//2023 IEEE/CVF International Conference on Computer Vision (ICCV). New York: IEEE, 2023: 19773-19786. DOI: 10.1109/ICCV51070.2023.01816.

[12] WANG R S, LEI T, CUI R X, *et al*. Medical Image Segmentation Using Deep Learning: A Survey[J]. *IET Image Process*, 2022, **16**(5): 1243-1267. DOI: 10.1049/ipr2.12419.

[13] YANG R H, SRIVASTAVA P, MANDT S. Diffusion Probabilistic Modeling for Video Generation[J]. *Entropy*, 2023, **25**(10): 1469. DOI: 10.3390/e25101469.

[14] WANG Y N, WU J M, FURUMAI K, *et al*. VAE-based Adversarial Multimodal Domain Transfer for Video-level Sentiment Analysis[J]. *IEEE Access*, 2022, **10**: 51315-

51324. DOI: 10.1109/ACCESS.2022.3174215.

[15] GU A, DAO T. Mamba: Linear-time Sequence Modeling with Selective State Spaces[EB/OL]. (2023−12−01)[2024−06−08]. https://doi.org/10.48550/arXiv.2312.00752.

[16] ZHU L H, LIAO B C, ZHANG Q, *et al*. Vision Mamba: Efficient Visual Representation Learning with Bidirectional State Space Model[EB/OL]. (2024−03−14)[2024−06−08]. https://doi.org/10.48550/arXiv.2401.09417.

[17] QIAO Y Y, YU Z, GUO L T, *et al*. VL-mamba: Exploring State Space Models for Multimodal Learning[EB/OL]. (2024−03−20) [2024−06−08]. https://doi. org/10.48550/arXiv.2403.13600.

[18] ZHAO H, ZHANG M, ZHAO W, *et al*. Cobra: Extending Mamba to Multi-modal Large Language Model for Efficient Inference[EB/OL]. (2024−06−05) [2024−06−08]. https://doi.org/10.48550/arXiv.2403.14520.

[19] LIEBER O, LENZ B, BATA H, *et al*. Jamba: A Hybrid Transformer-mamba Language Model[EB/OL]. (2024−03−28) [2024−06−08]. https://doi. org/10.48550/arXiv.2403.19887.

[20] CHEFER H, ALALUF Y, VINKER Y, *et al*. Attend-and-excite: Attention-based Semantic Guidance for Text-to-image Diffusion Models[J]. *ACM Trans Graph*, 2023, **42**(4): 1−10. DOI: 10.1145/3592116.

[21] YANG L, ZHANG Z L, SONG Y, *et al*. Diffusion Models: A Comprehensive Survey of Methods and Applications[J]. *ACM Comput Surv*, 2024, **56**(4): 1−39. DOI: 10.1145/3626235.