



(12) 发明专利

(10) 授权公告号 CN 115631103 B

(45) 授权公告日 2023.09.05

(21) 申请号 202211268479.8

G06T 5/50 (2006.01)

(22) 申请日 2022.10.17

G06V 10/82 (2022.01)

(65) 同一申请的已公布的文献号

G06N 3/04 (2023.01)

申请公布号 CN 115631103 A

G06N 3/08 (2023.01)

(43) 申请公布日 2023.01.20

(56) 对比文件

(73) 专利权人 北京百度网讯科技有限公司

CN 113112536 A, 2021.07.13

地址 100085 北京市海淀区上地十街10号

CN 114820398 A, 2022.07.29

百度大厦二层

CN 113822289 A, 2021.12.21

(72) 发明人 冯智达 张振宇 余欣彤 李岚欣

US 2022036516 A1, 2022.02.03

方晔玮 陈徐屹 刘佳祥 尹维冲

US 2020372611 A1, 2020.11.26

冯仕堃 孙宇

CN 112927159 A, 2021.06.08

(74) 专利代理机构 北京清亦华知识产权代理事

CN 110163827 A, 2019.08.23

务所(普通合伙) 11201

CN 110288547 A, 2019.09.27

专利代理师 张润

CN 113538281 A, 2021.10.22

CN 113781356 A, 2021.12.10

审查员 赵翠翠

(51) Int. Cl.

G06T 5/00 (2006.01)

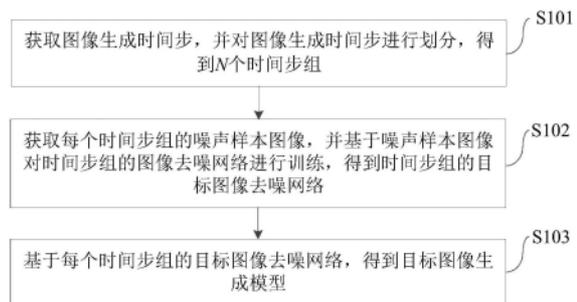
权利要求书3页 说明书12页 附图4页

(54) 发明名称

图像生成模型的训练方法和装置、图像生成方法及装置

(57) 摘要

本公开公开了一种图像生成模型的训练方法和装置,以及图像生成方法及装置,涉及人工智能技术领域,进一步涉及图像处理技术领域。具体实现方案为:获取图像生成时间步,并对图像生成时间步进行划分,得到N个时间步组,获取每个时间步组的噪声样本图像,并基于噪声样本图像对时间步组的图像去噪网络进行训练,得到时间步组的目标图像去噪网络,基于每个时间步组的目标图像去噪网络,得到目标图像生成模型。本公开实施例中,通过对每个时间步组分别进行图像去噪模型的训练,不仅可以考虑较近时间步的相近性,同时兼顾了较远时间步的差异性,能够在不增加模型训练的计算量的前提下,并减少不同网络任务之间的干扰,从而可以实现大幅提升模型精度。



1. 一种图像生成模型的训练方法,所述方法包括:

获取图像生成时间步,并对所述图像生成时间步进行划分,得到N个时间步组,其中,所述N为正整数且 $N \geq 2$ ,每个时间步组内的时间步在时序上连续,相邻的时间步组间时间步在时序也连续;

获取初始样本图像;对所述初始样本图像进行不同程度的加噪处理,得到不同的所述时间步组的噪声样本图像,并基于所述噪声样本图像对所述时间步组的图像去噪网络进行训练,得到所述时间步组的目标图像去噪网络;其中,所述时间步组内的时间步的数值与带噪数量成正相关;

基于每个所述时间步组的目标图像去噪网络,得到目标图像生成模型;

其中,所述方法还包括:

确定所述时间步组在所述图像生成时间步中的时序;

确定所述N个时间步组中所述时序为目标时序的第一时间步组;

确定所述第一时间步组的图像去噪网络与剩余的第二时间步组的图像去噪网络的类型不同。

2. 根据权利要求1所述的方法,其中,所述基于每个所述时间步组的目标图像去噪网络,得到目标图像生成模型,包括:

对所有的所述目标图像去噪网络在时间步维度上进行串联,得到所述目标图像生成模型。

3. 根据权利要求2所述的方法,其中,所述对所有的所述目标图像去噪网络在时间步维度上进行串联,得到所述目标图像生成模型,包括:

确定所述时间步组在所述图像生成时间步中的时序;

根据每个所述时间步组的时序,确定所述时间步组的目标图像去噪网络在所述目标图像生成模型中的串联顺序;

按照所述串联顺序对所述目标图像去噪网络进行串联,得到所述目标图像生成模型。

4. 根据权利要求1所述的方法,其中,所述对所述初始样本图像进行不同程度的加噪处理,得到不同的所述时间步组的噪声样本图像,包括:

确定所述时间步组在所述图像生成时间步中的时序;

根据所述时间步组的所述时序,确定所述初始样本图像的加噪数据;

基于所述加噪数据对所述初始样本图像进行加噪处理,得到所述时间步组的噪声样本图像。

5. 根据权利要求1所述的方法,其中,所述对所述初始样本图像进行不同程度的加噪处理,得到不同的所述时间步组的噪声样本图像,包括:

针对时间步组i,获取时间步组i-1的噪声样本图像;

获取预设的噪声增量数据;

按照所述噪声增量数据,对所述时间步组i-1的噪声样本图像进行噪声增强,得到所述时间步组i的噪声样本图像,其中, $1 \leq i \leq N$ 。

6. 根据权利要求1-5中任一项所述的方法,其中,时间上连续的M个时间步组的所述噪声样本图像相同,其中,所述M为正整数且 $2 \leq M \leq N$ 。

7. 根据权利要求1-5中任一项所述的方法,其中,所述N个时间步组中至少部分时间步

组对应的图像去噪网络的类型不同;或者,所述N个时间步组的图像去噪网络的类型相同。

8. 根据权利要求1-5中任一项所述的方法,其中,所述对所述图像生成时间步进行划分,得到N个时间步组,包括:

确定所述图像生成时间步的数量,并按照所述图像生成时间步的数量,确定所述N;

根据所述N和每个时间步的时序,确定每个所述时间步组所包括的时间步,其中,每个时间步组内的时间步在时序上连续。

9. 一种图像生成方法,所述方法包括:

获取噪声数据;

将所述噪声数据输入目标图像生成模型中,输出目标图像,其中,所述目标图像生成模型为采用权利要求1-8中任一项所述的训练方法训练得到的模型。

10. 根据权利要求9所述的方法,其中,所述获取噪声数据,包括:

随机采样高斯噪声,并将所述高斯噪声作为所述噪声数据;或者,

获取携带高斯噪声的图像,将所述携带高斯噪声的图像作为所述噪声数据。

11. 一种图像生成模型的训练装置,所述装置法包括:

获取模块,用于获取图像生成时间步,并对所述图像生成时间步进行划分,得到N个时间步组,其中,所述N为正整数且 $N \geq 2$ ,每个时间步组内的时间步在时序上连续,相邻的时间步组间时间步在时序也连续;

模型训练模块,用于获取初始样本图像;对所述初始样本图像进行不同程度的加噪处理,得到不同的所述时间步组的噪声样本图像,并基于所述噪声样本图像对所述时间步组的图像去噪网络进行训练,得到所述时间步组的目标图像去噪网络;其中,所述时间步组内的时间步的数值与带噪数量成正相关;

模型生成模块,用于基于每个所述时间步组的目标图像去噪网络,得到目标图像生成模型;

其中,所述模型训练模块,还用于:

确定所述时间步组在所述图像生成时间步中的时序;

确定所述N个时间步组中所述时序为目标时序的第一时间步组;

确定所述第一时间步组的图像去噪网络与剩余的第二时间步组的图像去噪网络的类型不同。

12. 根据权利要求11所述的装置,其中,所述模型生成模块,还用于:

对所有的所述目标图像去噪网络在时间步维度上进行串联,得到所述目标图像生成模型。

13. 根据权利要求12所述的装置,其中,所述模型生成模块,还用于:

确定所述时间步组在所述图像生成时间步中的时序;

根据每个所述时间步组的时序,确定所述时间步组的目标图像去噪网络在所述目标图像生成模型中的串联顺序;

按照所述串联顺序对所述目标图像去噪网络进行串联,得到所述目标图像生成模型。

14. 根据权利要求11所述的装置,其中,所述模型训练模块,还用于:

确定所述时间步组在所述图像生成时间步中的时序;

根据所述时间步组的所述时序,确定所述初始样本图像的加噪数据;

基于所述加噪数据对所述初始样本图像进行加噪处理,得到所述时间步组的噪声样本图像。

15. 根据权利要求11所述的装置,其中,所述模型训练模块,还用于:

针对时间步组 $i$ ,获取时间步组 $i-1$ 的噪声样本图像;

获取预设的噪声增量数据;

按照所述噪声增量数据,对所述时间步组 $i-1$ 的噪声样本图像进行噪声增强,得到所述时间步组 $i$ 的噪声样本图像,其中, $1 \leq i \leq N$ 。

16. 根据权利要求11-15中任一项所述的装置,其中,时间上连续的 $M$ 个时间步组的所述噪声样本图像相同,其中,所述 $M$ 为正整数且 $2 \leq M \leq N$ 。

17. 根据权利要求11-15中任一项所述的装置,其中,所述 $N$ 个时间步组中至少部分时间步组对应的图像去噪网络的类型不同;或者,所述 $N$ 个时间步组的图像去噪网络的类型相同。

18. 根据权利要求11-15中任一项所述的装置,其中,所述获取模块,还用于:

确定所述图像生成时间步的数量,并按照所述图像生成时间步的数量,确定所述 $N$ ;

根据所述 $N$ 和每个时间步的时序,确定每个所述时间步组所包括的时间步,其中,每个时间步组内的时间步在时序上连续。

19. 一种图像生成装置,所述装置包括:

获取模块,用于获取噪声数据;

图像生成模块,用于将所述噪声数据输入目标图像生成模型中,得到目标图像,其中,所述目标图像生成模型为采用权利要求1-8中任一项所述的训练装置训练得到的模型。

20. 根据权利要求19所述的装置,其中,所述获取模块,还用于:

随机采样高斯噪声,并将所述高斯噪声作为所述噪声数据;或者,

获取携带高斯噪声的图像,将所述携带高斯噪声的图像作为所述噪声数据。

21. 一种电子设备,包括:

至少一个处理器;以及

与所述至少一个处理器通信连接的存储器;其中,

所述存储器存储有可被所述至少一个处理器执行的指令,所述指令被所述至少一个处理器执行,以使所述至少一个处理器能够执行权利要求1-8中任一项所述的方法,或者以使所述至少一个处理器能够执行权利要求9或10所述的方法。

22. 一种存储有计算机指令的非瞬时计算机可读存储介质,其中,所述计算机指令用于使所述计算机执行根据权利要求1-8中任一项所述的方法,或者以使至少一个处理器能够执行权利要求9或10所述的方法。

23. 一种计算机程序产品,包括计算机程序,所述计算机程序在被处理器执行时实现根据权利要求1-8中任一项所述的方法,或者以使至少一个处理器能够执行权利要求9或10所述的方法。

## 图像生成模型的训练方法和装置、图像生成方法及装置

### 技术领域

[0001] 本公开涉及人工智能技术领域,进一步涉及图像技术领域。

### 背景技术

[0002] 不同的时间步(timestep)的网络任务往往存在差异,但是对于很大的时间步,网络模型的输入可以近似于纯高斯噪声,此时网络任务可以看作一个图像生成任务,即从噪声中生成一张含有语义的图像。而对于较小的时间步,网络模型的输入为带噪声的图像,网络任务往往是对图像的细节进行优化。相关技术中,网络模型对所有的时间步往往使用的相同的模型参数,忽略了各个时间步之间的差异,最终导致训练出的网络模型的性能较差。

### 发明内容

[0003] 本公开提供了一种图像生成模型的训练方法及装置,通过对图像生成时间步进行划分,为每个时间步组进行单独的图像去噪网络的训练,可以考虑各个时间步之间的差异,使得训练出的网络模型的性能较高。

[0004] 根据本公开的另一方面,提供了一种图像生成方法,可以通过训练出的目标图像生成模型,基于噪声数据生成较高质量的图像。

[0005] 根据本公开的另一方面,提供了一种图像生成模型的装置。

[0006] 根据本公开的另一方面,提供了一种图像生成装置。

[0007] 根据本公开的另一方面,提供了一种电子设备。

[0008] 根据本公开的另一方面,提供了一种非瞬时计算机可读存储介质。

[0009] 根据本公开的另一方面,提供了一种计算机程序产品。

[0010] 为达上述目的,本公开第一方面实施例提出了一种图像生成模型的训练方法,包括:

[0011] 获取图像生成时间步,并对所述图像生成时间步进行划分,得到N个时间步组,其中,所述N为正整数且 $N \geq 2$ ;

[0012] 获取每个所述时间步组的噪声样本图像,并基于所述噪声样本图像对所述时间步组的图像去噪网络进行训练,得到所述时间步组的目标图像去噪网络;

[0013] 基于每个所述时间步组的目标图像去噪网络,得到目标图像生成模型。

[0014] 为达上述目的,本公开第二方面实施例提出了一种图像生成方法,包括:

[0015] 获取噪声数据;

[0016] 将所述噪声数据输入目标图像生成模型中,输出目标图像,其中,所述目标图像生成模型为采用第一方面实施例提出了一种图像生成模型的训练方法训练出的模型。

[0017] 为达上述目的,本公开第三方面实施例提出了一种图像生成模型的训练装置,包括:

[0018] 获取模块,用于获取图像生成时间步,并对所述图像生成时间步进行划分,得到N

个时间步组,其中,所述N为正整数且 $N \geq 2$ ;

[0019] 模型训练模块,用于获取每个所述时间步组的噪声样本图像,并基于所述噪声样本图像对所述时间步组的图像去噪网络进行训练,得到所述时间步组的目标图像去噪网络;

[0020] 模型生成模块,用于基于每个所述时间步组的目标图像去噪网络,得到目标图像生成模型。

[0021] 为达上述目的,本公开第四方面实施例提出了一种图像生成装置,包括:

[0022] 获取模块,用于获取噪声数据;

[0023] 图像生成模块,用于将所述噪声数据输入目标图像生成模型中,得到目标图像,其中,所述目标图像生成模型为采用第一方面实施例提出了一种图像生成模型的训练方法训练出的模型。

[0024] 为达上述目的,本公开第五方面实施例提出了一种电子设备,包括存储器、处理器。所述存储器存储有可被所述至少一个处理器执行的指令,所述指令被所述至少一个处理器执行,以实现如本公开第一方面实施例所述的方法,或者用于实现如本公开第二方面实施例所述的方法。

[0025] 为达上述目的,本公开第六方面实施例提出了一种存储有计算机指令的非瞬时计算机可读存储介质,其中,所述计算机指令用于实现如本公开第一方面实施例所述的方法,或者用于实现如本公开第二方面实施例所述的方法。

[0026] 为达上述目的,本公开第七方面实施例提出了一种计算机程序产品,包括计算机程序,所述计算机程序在被处理器执行时以实现如本公开第一方面实施例所述的方法,或者用于实现如本公开第二方面实施例所述的方法。

[0027] 应当理解,本部分所描述的内容并非旨在标识本公开的实施例的关键或重要特征,也不用于限制本公开的范围。本公开的其它特征将通过以下的说明书而变得容易理解。

## 附图说明

[0028] 附图用于更好地理解本方案,不构成对本公开的限定。其中:

[0029] 图1是根据本公开实施例一种图像生成模型的训练方法的示意图;

[0030] 图2是根据本公开实施例另一种图像生成模型的训练方法的示意图;

[0031] 图3是根据本公开实施例另一种图像生成模型的训练方法的示意图;

[0032] 图4是根据本公开实施例一种图像生成模型的结构示意图;

[0033] 图5是根据本公开实施例一种图像生成方法的流程示意图;

[0034] 图6是根据本公开实施例一种图像生成模型的训练装置的示意图;

[0035] 图7是根据本公开实施例一种图像生成装置的示意图;

[0036] 图8是根据本公开实施例一种电子设备的结构图。

## 具体实施方式

[0037] 以下结合附图对本公开的示范性实施例做出说明,其中包括本公开实施例的各种细节以助于理解,应当将它们认为仅仅是示范性的。因此,本领域普通技术人员应当认识到,可以对这里描述的实施例做出各种改变和修改,而不会背离本公开的范围和精神。同

样,为了清楚和简明,以下的描述中省略了对公知功能和结构的描述。

[0038] 这里将详细地对示例性实施例进行说明,其示例表示在附图中。下面的描述涉及附图时,除非另有表示,不同附图中的相同数字表示相同或相似的要素。以下示例性实施例中所述的实施方式并不代表与本公开实施例相一致的所有实施方式。相反,它们仅是与如所附权利要求书中所详述的、本公开实施例的一些方面相一致的装置和方法的例子。

[0039] 在本公开实施例使用的术语是仅仅出于描述特定实施例的目的,而非旨在限制本公开实施例。在本公开实施例和所附权利要求书中所使用的单数形式的“一种”和“该”也旨在包括多数形式,除非上下文清楚地表示其他含义。还应当理解,本文中使用的术语“和/或”是指并包含一个或多个相关联的列出项目的任何或所有可能组合。

[0040] 应当理解,尽管在本公开实施例可能采用术语第一、第二、第三等来描述各种信息,但这些信息不应限于这些术语。这些术语仅用来将同一类型的信息彼此区分开。例如,在不脱离本公开实施例范围的情况下,第一信息也可以被称为第二信息,类似地,第二信息也可以被称为第一信息。取决于语境,如在此所使用的词语“如果”及“若”可以被解释成为“在……时”或“当……时”或“响应于确定”。

[0041] 图像处理(Image Processing),用计算机对图像进行分析,以达到所需结果的技术。又称影像处理。图像处理一般指数字图像处理。数字图像是指用工业相机、摄像机、扫描仪等设备经过拍摄得到的一个大的二维数组,该数组的元素称为像素,其值称为灰度值。图像处理技术一般包括图像压缩,增强和复原,匹配、描述和识别3个部分。

[0042] 深度学习(Deep Learning,简称DL),是机器学习(Machine Learning,简称ML)领域中一个新的研究方向,它被引入机器学习使其更接近于最初的目标——人工智能。深度学习是学习样本数据的内在律和表示层次,这些学习过程中获得的信息对诸如文字,图像和声音等数据的解释有很大的帮助。它的最终目标是让机器能够像人一样具有分析学习能力,能够识别文字、图像和声音等数据。深度学习是一个复杂的机器学习算法,在语音和图像识别方面取得的效果,远远超过先前相关技术。

[0043] 计算机视觉(Computer Vision),是一门研究如何使机器“看”的科学,更进一步的说,就是指用摄影机和电脑代替人眼对目标进行识别、跟踪和测量等机器视觉,并进一步做图形处理,使电脑处理成为更适合人眼观察或传送给仪器检测的图像。作为一个科学学科,计算机视觉研究相关的理论和技术,试图建立能够从图像或者多维数据中获取‘信息’的人工智能系统。这里所指的信息指Shannon定义的,可以用来帮助做一个“决定”的信息。因为感知可以看作是从感官信号中提取信息,所以计算机视觉也可以看作是研究如何使人工系统从图像或多维数据中“感知”的科学。

[0044] 人工智能(Artificial Intelligence,简称AI),是研究使计算机来模拟人生的某些思维过程和智能行为(如学习、推理、思考、规划等)的学科,既有硬件层面的技术,也有软件层面的技术。人工智能硬件技术一般包括计算机视觉技术、语音识别技术、自然语言处理技术以及及其学习/深度学习、大数据处理技术、知识图谱技术等几大方面。

[0045] 图1为本公开实施例提供的一种图像生成模型的训练方法的流程示意图。如图1所示,该方法包括但不限于以下步骤:

[0046] S101,获取图像生成时间步,并对图像生成时间步进行划分,得到N个时间步组。

[0047] 其中,所述N为正整数且 $N \geq 2$ 。

[0048] 图像生成时间步可以理解为生成一个图像所需的步数和每步的时序,例如,图像生成所需的步数可以为1000步,1000步按照时序先后排列可以称为一个图像生成时间步,即从第1步至第1000步可以称为一个图像生成时间步。

[0049] 在一些实现中,图像生成时间步可以基于模型的精度和模型的运算量来确定。可选地,图像生成时间步可以在预先配置的训练参数中获取,也可以在模型训练开始时进行配置。

[0050] 由于相近的时间步的任务比较类似,本公开实施例中,为了减少神经网络的数量,可以对图像生成时间步进行划分,得到N个时间步组,为每个时间步组训练一个神经网络,无需为每个时间步都训练一个神经网络,从而可以减少神经网络的数量,降低训练时所需的运算量和资源。

[0051] 在一些实现中,可以对图像生成时间步进行均匀划分,得到N个时间步组。例如,1000步可以均匀划分成20个时间步组,每一组包括50个时间步。

[0052] 在一些实现中,可以对图像生成时间步进行非均匀划分,得到N个时间步组。例如,1000步可以非均匀划分成20个时间步组,可选地,可以按照时间步组的序号逐渐增多每个时间步组内的时间步数量。可选地,可以随机确定每个时间步组内的时间步数量。比方,第5个时间步组内包括20个时间步,第10个时间步组内包括100个时间步,第15个时间步组内包括20个时间步。可以理解的时,不同时间步组内包括的时间步的数量可以相同或不同。

[0053] 需要说明的是,每个时间步组内的时间步在时序上连续,相邻的时间步组间时间步在时序也连续,即第i个时间步组与第i+1个时间步组所包括的时间步在时序上连续,也就是说,第i个时间步组的最后一个时间步为第10步,则第i+1个时间步组内的首个时间步为第11步。

[0054] S102,获取每个时间步组的噪声样本图像,并基于噪声样本图像对时间步组的图像去噪网络进行训练,得到时间步组的目标图像去噪网络。

[0055] 本公开实施例中,可以为每个时间步组设置一个图像去噪网络,通过训练每个时间步组的图像去噪网络,由于时间步组在时序上连续,在得到每个时间步组训练后的目标图像去噪网络后,可以对目标图像去噪网络在时间步维度上进行组合,得到最终的目标图像生成模型。

[0056] 在一些实现中,可以为每个时间步组设置对应的噪声样本图像。可选地,可以对同样的初始样本图像进行加噪处理,不同的时间步组对应的加噪数据可以不同,进而得到不同的时间步组的噪声样本图像。

[0057] 可选地,N个时间步组中时序连续的M个时间步组,可以采用相同的噪声样本图像。例如,时间步组t、时间步组t-1和时间步组t-2,可以采用相同的噪声样本图像。在相近的时间步组上采用相同的噪声样本图像进行训练,不仅可以降低采集噪声样本图像所需的资源和时间,而且相近的时间步组的任务差异较小,采用相同的噪声样本图像可以维持相近的时间步组的图像去噪网络的训练相近性,提高最终图像生成的精度。可以理解的是,时间步组t、时间步组t-1、时间步组t-2,可以采用相同的第一噪声样本图像;时间步组k+1、时间步组k、时间步组k-1、时间步组k-2,可以采用相同的第二噪声样本图像。

[0058] 可选地,图像去噪网络可以采用去噪扩散网络。

[0059] S103,基于每个时间步组的目标图像去噪网络,得到目标图像生成模型。

[0060] 在模型推理阶段的图像生成时间步是不变的,本申请实施例中,可以按照顺序对所有的目标图像去噪模型进行串联,得到一个完整的目标图像生成模型。

[0061] 在得到每个时间步组训练后的目标图像去噪网络后,可以对所有的目标图像去噪网络在时间步维度上进行组合,得到最终的目标图像生成模型。

[0062] 在一些实现中,可以基于时间步组的序号,确定目标图像去噪网络的组合顺序,按照顺序对所有的目标图像去噪网络在时间步维度上进行组合,得到最终的目标图像生成模型。

[0063] 在一些实现中,可以确定时间步组在图像生成时间步的时序,基于该时间步组的时序,确定目标图像去噪网络的组合顺序,按照顺序对所有的目标图像去噪网络在时间步维度上进行组合,得到最终的目标图像生成模型。

[0064] 本公开实施例中,获取图像生成时间步,并对图像生成时间步进行划分,得到N个时间步组,获取每个时间步组的噪声样本图像,并基于噪声样本图像对时间步组的图像去噪网络进行训练,得到时间步组的目标图像去噪网络,基于每个时间步组的目标图像去噪网络,得到目标图像生成模型。本公开实施例中,通过对每个时间步组分别进行图像去噪模型的训练,使得图像生成模型的训练过程中,较近的时间步差异较小,可以作为一组并对对应训练一个图像去噪网络,无需对每个时间步进行训练,可以节省图像去噪网络的数量节省资源。而对于相距较远的时间步之间的差异,通过单独训练使得图像去噪网络考虑了时间步之间的差异,能够使得模型训练的过程更加合理,使得最终的目标图像生成模型在基于噪声生成图像时,能够得到质量较高的图像。而且通过单独训练能够让不同的图像去噪网络去专门学习各自的网络任务,可以减少不同类型的网络任务之间的干扰,提高推理阶段混合后整体图像生成模型的性能。

[0065] 图2为本公开实施例提供的一种图像生成模型的训练方法的流程示意图。如图2所示,该方法包括但不限于以下步骤:

[0066] S201,获取图像生成时间步,并对图像生成时间步进行划分,得到N个时间步组。

[0067] 在一些实现中,确定图像生成时间步的数量,并按照图像生成时间步的数量,确定N的取值,例如,不同的数量可以对应不同的取值。进一步地,根据N和每个时间步的时序(步数编号),确定每个时间步组所包括的时间步,其中,每个时间步组内的时间步在时序上连续。

[0068] 示例性说明,1000步可以均匀地划分成20个时间步组,其中,第1个时间步组可以包括第1步~第50步,第2个时间步组可以包括第51步~第100步,以此类推,第20个时间步组可以包括第951步~第1000步。

[0069] S202,获取初始样本图像,并对初始样本图像进行不同程度的加噪处理,得到不同的时间步组的噪声样本图像。

[0070] 其中,时间步组内的时间步的数值与带噪数量成正相关。

[0071] 在一些实现中,可以从图像库中随机选取部分图像,作为初始样本图像。

[0072] 进一步地,对初始样本图像进行加噪处理,以获取到噪声样本图像。本公开实施例中,在初始样本图像中所加的噪声为高斯噪声。通过对初始样本图像进行加噪处理,能够得到每个时间步组的噪声样本图像,从而可以实现单独训练多个图像去噪网络的目的。

[0073] 为了使得N个时间步组可以得到各自的噪声样本图像,可以对初始样本图像进行

不同程度的加噪处理,其中,时间步组在图像生成时间步中的时序越靠后,对应的加噪处理的程度越高,也就是说,时间步组内时间步的数值与图像带噪的数量成正相关。

[0074] 示例性说明,第1个时间步组可以包括第1步~第50步,第2个时间步组可以包括第51步~第100步,……以此类推,第19个时间步组可以包括第901步~第950步,第20个时间步组可以包括第951步~第1000步。由于时间步组*i*内的时间步的数值大于时间步组*i-1*的时间步的数值,本公开实施例中,时间步组*i*的噪声样本图像上的噪声数量,大于时间步组*i-1*的噪声样本图像上的噪声数量。

[0075] 作为一种可能的实现方式,可以确定时间步组在图像生成时间步中的时序,进一步地,针对每个时间步组,根据时间步组的时序,确定初始样本图像的加噪数据,基于加噪数据对初始样本图像进行加噪处理,得到时间步组的噪声样本图像,其中时序越靠前,相应地噪声越少,时序越靠后,相应地噪声数据越大。可选地,可以基于时间步组的序号,确定该时间步组在图像生成时间步中的时序,例如,序号越大时序越靠后。

[0076] 在该实现方式中,为每个时间步组进行随机高斯噪声的生成,以到每个时间步组的噪声样本图像,从而可以满足基于不同的噪声样本图像对不同时间步组进行训练的需求。

[0077] 作为另一种可能的实现方式,针对时间步组*i*,获取时间步组*i-1*的噪声样本图像,获取预设的噪声增量数据,按照噪声增量数据,对时间步组*i-1*的噪声样本图像进行噪声增强,得到时间步组*i*的噪声样本图像,其中, $1 \leq i \leq N$ 。也就是说,从首个时间步组开始,在前一个时间步组的基础上,对前一个时间步组的噪声样本图像进行噪声增强,得到当前时间步组的噪声样本图像。

[0078] 在该实现方式中,在首个时间步组的基础上,通过一个噪声增量数据,可以获取到后续每个时间步组的噪声样本图像,无需每个都进行随机高斯噪声的生成,提高噪声样本图像的获取效率,而且可以满足单独训练每个时间步组的图像去噪网络的需求。

[0079] S203,基于噪声样本图像对时间步组的图像去噪网络进行训练,得到时间步组的目标图像去噪网络。

[0080] S204,基于每个时间步组的目标图像去噪网络,得到目标图像生成模型。

[0081] 关于步骤S204~S205的具体介绍,可参见上述实施例中相关步骤的记载,此处不再赘述。

[0082] 本公开实施例中,通过对每个时间步组分别进行图像去噪模型的训练,使得图像生成模型的训练过程中,较近的时间步差异较小,可以作为一组并对应训练一个图像去噪网络,无需对每个时间步进行训练,可以节省图像去噪网络的数量节省资源。而对于相距较远的时间步之间的差异,通过单独训练使得图像去噪网络考虑了时间步之间的差异,能够使得模型训练的过程更加合理,使得最终的目标图像生成模型在基于噪声生成图像时,能够得到质量较高的图像。

[0083] 图3为本公开实施例提供的一种图像生成模型的训练方法的流程示意图。如图3所示,该方法包括但不限于以下步骤:

[0084] S301,获取图像生成时间步,并对图像生成时间步进行划分,得到*N*个时间步组。

[0085] S302,获取每个时间步组的噪声样本图像,并基于噪声样本图像对时间步组的图像去噪网络进行训练,得到时间步组的目标图像去噪网络。

[0086] 关于步骤S301~S302的具体介绍,可参见上述实施例中相关步骤的记载,此处不再赘述。

[0087] S303,确定时间步组在图像生成时间步中的时序。

[0088] 可选地,可以基于时间步组的序号,确定该时间步组在图像生成时间步中的时序,例如,序号越大时序越靠后,例如,时间步组的序号为*i*,则该时间步组在图像生成时间步中的时序即为第*i*个执行。

[0089] 可选地,可以基于时间步组内的时间步的数值,确定该时间步组在图像生成时间步中的时序,例如,数值越大时序越靠后。

[0090] S304,根据每个时间步组的时序,确定时间步组的目标图像去噪网络在目标图像生成模型中的串联顺序。

[0091] S305,按照串联顺序对目标图像去噪网络进行串联,得到目标图像生成模型。

[0092] 示例性说明,1000步被均匀地划分为20个时间步组,时间步组1、时间步组2、时间步组3,……,时间步组19和时间步组20。本公开实施例中,可以确定时间步组20对应的目标图像去噪网络20为首个网络,时间步组19对应的目标图像去噪网络19为第二个网络,时间步组18对应的目标图像去噪网络18为第三个网络,……,时间步组2对应的目标图像去噪网络2为第19个网络、时间步组1对应的目标图像去噪网络1为最后一个网络即第20个网络。

[0093] 本公开实施例中,可以将目标图像去噪网络20与目标图像去噪网络19连接,目标图像去噪网络19与目标图像去噪网络18连接,目标图像去噪网络18与目标图像去噪网络17连接,……,目标图像去噪网络2与目标图像去噪网络1连接,最终得到目标图像生成模型。

[0094] 在一些实现中,*N*个时间步组中至少部分时间步组对应的图像去噪网络的类型不同;或者,*N*个时间步组的图像去噪网络的类型相同。例如,可以包括两个网络单元,两个网络单元连接,第一网络单元进行下采样,第二网络单元进行上采样。

[0095] 可选地,根据时间步组在图像生成时间步中的时序,进一步地,可以确定*N*个时间步组中所述时序为目标时序的第一时间步组,确定第一时间步组的图像去噪网络的类型与剩余的第二天步组的图像去噪网络的类型不同。例如,目标时序可以为首个执行的时间步组,或者最后一个执行的时间步组,或者中间特定的一个或多个时间步组。可以将这些目标时序的一个或多个时间步组称为第一时间步组。本公开实施例中,可以构建第一时间步组的图像去噪网络的结构或类型,与剩余的第二天步组的图像去噪网络的结构或类型不同。通过设置不同的图像去噪网络,能够丰富网络结构或类型,使得图像去噪过程具有更高的兼容性,可以更好地实现图像去噪,提高图像生成的效率。

[0096] 本公开实施例中,通过对每个时间步组分别进行图像去噪模型的训练,而在模型推理阶段的图像生成时间步是不变的,从而可以基于时间步组的时序/编号等,对所有的目标图像去噪模型按序进行串联,得到一个完整的目标图像生成,通过本申请实施例提供的训练方式,不仅可以考虑较近时间步的相近性,同时兼顾了较远时间步的差异性,即可以减少不同类型的网络任务之间的干扰,从而可以在不增加模型训练的计算量的前提下,实现大幅提升模型精度和容量的目的。

[0097] 图4为本公开实施例提供的一种图像生成模型的训练方法的流程示意图。如图4所示,时间步组*N*内包括的时间步为 $x_{T_n} \sim x_{T_{n-1}}$ ,该时间步组*N*对应的图像去噪网络(Denoising U-Net No.*N*),时间步组*N*-1内包括的时间步为 $x_{T_{n-1}} \sim x_{T_{n-2}}$ ,时间步组*N*-1对应的图像去噪网

络(Denoising U-NetNo.N-1);……;时间步组t内包括的时间步为 $x_{T_t} \sim x_{T_{t-1}}$ ,时间步组t对应的图像去噪网络(Denoising U-Net No.t);……;时间步组1内包括的时间步为 $x_0 \sim x_{T_1}$ ,时间步组1对应的图像去噪网络(Denoising U-Net No.1)。如图4所示,每个Denoising U-Net可以包括两个网络单元,两个网络单元连接,前一个网络单元对输入的噪声样本图像进行下采样,后一个网络单元对下采样后的图像进行上采样,以恢复到原始图像大小。

[0098] 在推理阶段,可以从Denoising U-NetNo.N→Denoising U-NetNo.N-1→……→Denoising U-Net No.t→……→Denoising U-Net No.1。需要说明的是,在训练阶段,各个Denoising U-Net之间单独训练,并不存在箭头所指的训练次序。

[0099] 图5为本公开实施例提供的一种图像生成方法的流程示意图。如图5所示,该方法包括但不限于以下步骤:

[0100] S501,获取噪声数据。

[0101] 在一些实现中,可以随机采样高斯噪声,并将该随机采样的高斯噪声确定为用于生成图像的噪声数据。

[0102] 在一些实现中,可以获取携带高斯噪声的图像,将该携带高斯噪声的图像确定为噪声数据。

[0103] S502,将噪声数据输入目标图像生成模型中,输出目标图像。

[0104] 需要说明的是,本公开实施例中采用的目标图像生成模型为采用上述实施例提供的模型训练方法得到图像生成模型。具体过程可参见上述实施例中相关内容的记载,此处不再赘述。

[0105] 在噪声数据为随机采样的高斯噪声的情况下,目标图像生成模型可以通过上述输出一个图像。

[0106] 在噪声数据为携带高斯噪声图像的情况下,目标图像生成模型可以输出一个去除高斯噪声的原始图像。

[0107] 本公开实施例中,可以基于上述实施例训练出的目标图像生成模型,可以基于噪声数据生成一个所需要的图像,无需掌握复杂的绘画技巧,从而可以丰富图像资源,有利于提高优化相关应用,提高应用的粘度。在带噪图像的场景下可以能够提高图像还原的精度。

[0108] 图6为本公开实施例提供的一种图像生成模型的训练装置的结构示意图。如图6所示,该装置包括但不限于以下模块:获取模块601、模型训练模块602和模型生成模块603。

[0109] 其中,获取模块601,用于获取图像生成时间步,并对所述图像生成时间步进行划分,得到N个时间步组,其中,所述N为正整数且 $N \geq 2$ ;

[0110] 模型训练模块602,用于获取每个所述时间步组的噪声样本图像,并基于所述噪声样本图像对所述时间步组的图像去噪网络进行训练,得到所述时间步组的目标图像去噪网络;

[0111] 模型生成模块603,用于基于每个所述时间步组的目标图像去噪网络,得到目标图像生成模型。

[0112] 在一些实现中,模型生成模块603,还用于:

[0113] 对所有的所述目标图像去噪网络在时间步维度上进行串联,得到所述目标图像生成模型。

[0114] 在一些实现中,模型生成模块603,还用于:

- [0115] 确定所述时间步组在所述图像生成时间步中的时序；
- [0116] 根据每个所述时间步组的时序，确定所述时间步组的目标图像去噪网络在所述目标图像生成模型中的串联顺序；
- [0117] 按照所述串联顺序对所述目标图像去噪网络进行串联，得到所述目标图像生成模型。
- [0118] 在一些实现中，模型训练模块602，还用于：
- [0119] 获取初始样本图像；
- [0120] 对所述初始样本图像进行不同程度的加噪处理，得到不同的所述时间步组的噪声样本图像，其中，所述时间步组内的时间步的数值与带噪数量成正相关。
- [0121] 在一些实现中，模型训练模块602，还用于：
- [0122] 确定所述时间步组在所述图像生成时间步中的时序；
- [0123] 根据所述时间步组的所述时序，确定所述初始样本图像的加噪数据；
- [0124] 基于所述加噪数据对所述初始样本图像进行加噪处理，得到所述时间步组的噪声样本图像。
- [0125] 在一些实现中，模型训练模块602，还用于：
- [0126] 针对时间步组 $i$ ，获取时间步组 $i-1$ 的噪声样本图像；
- [0127] 获取预设的噪声增量数据；
- [0128] 按照所述噪声增量数据，对所述时间步组 $i-1$ 的噪声样本图像进行噪声增强，得到所述时间步组 $i$ 的噪声样本图像，其中， $1 \leq i \leq N$ ；
- [0129] 在一些实现中，时间上连续的 $M$ 个时间步组的所述噪声样本图像相同，其中，所述 $M$ 为正整数且 $2 \leq M \leq N$ 。
- [0130] 在一些实现中， $N$ 个时间步组中至少部分时间步组对应的图像去噪网络的类型不同；或者，所述 $N$ 个时间步组的图像去噪网络的类型相同。
- [0131] 在一些实现中，模型训练模块602，还用于：
- [0132] 确定所述时间步组在所述图像生成时间步中的时序；
- [0133] 确定所述 $N$ 个时间步组中所述时序为目标时序的第一时间步组；
- [0134] 确定所述第一时间步组的图像去噪网络与剩余的第二时间步组的图像去噪网络的类型不同。
- [0135] 在一些实现中，获取模块601，还用于：
- [0136] 确定所述图像生成时间步的数量，并按照所述图像生成时间步的数量，确定所述 $N$ ；
- [0137] 根据所述 $N$ 和每个时间步的时序，确定每个所述时间步组所包括的时间步，其中，每个时间步组内的时间步在时序上连续。
- [0138] 本公开实施例中，通过对每个时间步组分别进行图像去噪模型的训练，使得图像生成模型的训练过程中，较近的时间步差异较小，可以作为一组并对应训练一个图像去噪网络，无需对每个时间步进行训练，可以节省图像去噪网络的数量节省资源。而对于相距较远的时间步之间的差异，通过单独训练使得图像去噪网络考虑了时间步之间的差异，能够使得模型训练的过程更加合理，使得最终的目标图像生成模型在基于噪声生成图像时，能够得到质量较高的图像。而且通过单独训练能够让不同的图像去噪网络去专门学习各自的

网络任务,可以减少不同类型的网络任务之间的干扰,提高推理阶段混合后整体图像生成模型的性能。

[0139] 图7为本公开实施例提供的一种图像生成装置的结构示意图。如图7所示,该装置包括但不限于以下模块:获取模块701和图像生成模块702。

[0140] 其中,获取模块701,用于获取噪声数据;

[0141] 图像生成模块702,用于将所述噪声数据输入目标图像生成模型中,得到目标图像,其中,所述目标图像生成模型为采用权利要求1-10中任一项所述的训练装置训练得到的模型。

[0142] 在一些实现中,获取模块701,还用于:

[0143] 随机采样高斯噪声,并将所述高斯噪声作为所述噪声数据;或者,

[0144] 获取携带高斯高斯噪声的图像,将所述携带高斯高斯噪声的图像作为所述噪声数据。

[0145] 本公开实施例中,可以基于上述实施例训练出的目标图像生成模型,可以基于噪声数据生成一个所需要的图像,无需掌握复杂的绘画技巧,从而可以丰富图像资源,有利于提高优化相关应用,提高应用的粘度。在带噪图像的场景下可以能够提高图像还原的精度。

[0146] 根据本公开的实施例,本公开还提供了一种电子设备、一种可读存储介质和一种计算机程序产品。

[0147] 图8示出了可以用来实施本公开的实施例的示例电子设备800的示意性框图。电子设备旨在表示各种形式的数字计算机,诸如,膝上型计算机、台式计算机、工作台、个人数字助理、服务器、刀片式服务器、大型计算机、和其它适合的计算机。电子设备还可以表示各种形式的移动装置,诸如,个人数字处理、蜂窝电话、智能电话、可穿戴设备和其它类似的计算装置。本文所示的部件、它们的连接和关系、以及它们的功能仅仅作为示例,并且不意在限制本文中描述的和/或者要求的本公开的实现。

[0148] 如图8所示,设备800包括计算单元801,其可以根据存储在只读存储器(ROM)802中的计算机程序或者从存储单元808加载到随机访问存储器(RAM)803中的计算机程序,来执行各种适当的动作和处理。在RAM 803中,还可存储设备800操作所需的各种程序和数据。计算单元801、ROM 802以及RAM 803通过总线804彼此相连。输入/输出(I/O)接口805也连接至总线804。

[0149] 设备800中的多个部件连接至I/O接口805,包括:输入单元806,例如键盘、鼠标等;输出单元807,例如各种类型的显示器、扬声器等;存储单元808,例如磁盘、光盘等;以及通信单元809,例如网卡、调制解调器、无线通信收发机等。通信单元809允许设备800通过诸如因特网的计算机网络和/或各种电信网络与其他设备交换信息/数据。

[0150] 计算单元801可以是各种具有处理和计算能力的通用和/或专用处理组件。计算单元801的一些示例包括但不限于中央处理单元(CPU)、图形处理单元(GPU)、各种专用的人工智能(AI)计算芯片、各种运行机器学习模型算法的计算单元、数字信号处理器(DSP)、以及任何适当的处理器、控制器、微控制器等。计算单元801执行上文所描述的各个方法和处理,例如方法XXX。例如,在一些实施例中,方法XXX可被实现为计算机软件程序,其被有形地包含于机器可读介质,例如存储单元808。在一些实施例中,计算机程序的部分或者全部可以经由ROM 802和/或通信单元809而被载入和/或安装到设备800上。当计算机程序加载到RAM

803并由计算单元801执行时,可以执行上文描述的方法XXX的一个或多个步骤。备选地,在其他实施例中,计算单元801可以通过其他任何适当的方式(例如,借助于固件)而被配置为执行方法XXX。

[0151] 本文中以上描述的系统和技术各种实施方式可以在数字电子电路系统、集成电路系统、场可编程门阵列(FPGA)、专用集成电路(ASIC)、专用标准产品(ASSP)、芯片上系统的系统(SOC)、负载可编程逻辑设备(CPLD)、计算机硬件、固件、软件、和/或它们的组合中实现。这些各种实施方式可以包括:实施在一个或者多个计算机程序中,该一个或者多个计算机程序可在包括至少一个可编程处理器的可编程系统上执行和/或解释,该可编程处理器可以是专用或者通用可编程处理器,可以从存储系统、至少一个输入装置、和至少一个输出装置接收数据和指令,并且将数据和指令传输至该存储系统、该至少一个输入装置、和该至少一个输出装置。

[0152] 用于实施本公开的方法的程序代码可以采用一个或多个编程语言的任何组合来编写。这些程序代码可以提供给通用计算机、专用计算机或其他可编程数据处理装置的处理器或控制器,使得程序代码当由处理器或控制器执行时使流程图和/或框图中所规定的功能/操作被实施。程序代码可以完全在机器上执行、部分地在机器上执行,作为独立软件包部分地在机器上执行且部分地在远程机器上执行或完全在远程机器或服务器上执行。

[0153] 在本公开的上下文中,机器可读介质可以是有形的介质,其可以包含或存储以供指令执行系统、装置或设备使用或与指令执行系统、装置或设备结合地使用的程序。机器可读介质可以是机器可读信号介质或机器可读储存介质。机器可读介质可以包括但不限于电子的、磁性的、光学的、电磁的、红外的、或半导体系统、装置或设备,或者上述内容的任何合适组合。机器可读存储介质的更具体示例会包括基于一个或多个线的电气连接、便携式计算机盘、硬盘、随机存取存储器(RAM)、只读存储器(ROM)、可擦除可编程只读存储器(EPROM或快闪存储器)、光纤、便捷式紧凑盘只读存储器(CD-ROM)、光学储存设备、磁储存设备、或上述内容的任何合适组合。

[0154] 为了提供与用户的交互,可以在计算机上实施此处描述的系统和技术,该计算机具有:用于向用户显示信息的显示装置(例如,CRT(阴极射线管)或者LCD(液晶显示器)监视器);以及键盘和指向装置(例如,鼠标或者轨迹球),用户可以通过该键盘和该指向装置来将输入提供给计算机。其它种类的装置还可以用于提供与用户的交互;例如,提供给用户的反馈可以是任何形式的传感反馈(例如,视觉反馈、听觉反馈、或者触觉反馈);并且可以用任何形式(包括声输入、语音输入或者、触觉输入)来接收来自用户的输入。

[0155] 可以将此处描述的系统和技术实施在包括后台部件的计算系统(例如,作为数据服务器)、或者包括中间件部件的计算系统(例如,应用服务器)、或者包括前端部件的计算系统(例如,具有图形用户界面或者网络浏览器的用户计算机,用户可以通过该图形用户界面或者该网络浏览器来与此处描述的系统和技术实施方式交互)、或者包括这种后台部件、中间件部件、或者前端部件的任何组合的计算系统中。可以通过任何形式或者介质的数字数据通信(例如,通信网络)来将系统的部件相互连接。通信网络的示例包括:局域网(LAN)、广域网(WAN)、互联网和区块链网络。

[0156] 计算机系统可以包括客户端和服务端。客户端和服务端一般远离彼此并且通常通过通信网络进行交互。通过在相应的计算机上运行并且彼此具有客户端-服务器关系的计

计算机程序来产生客户端和服务器的关系。服务端可以是云服务器,又称为云计算服务器或云主机,是云计算服务体系中的一项主机产品,以解决了传统物理主机与VPS服务(“Virtual Private Server”,或简称“VPS”)中,存在的管理难度大,业务扩展性弱的缺陷。服务器也可以为分布式系统的服务器,或者是结合区块链的服务器。

[0157] 应该理解,可以使用上面所示的各种形式的流程,重新排序、增加或删除步骤。例如,本发公开中记载的各步骤可以并行地执行也可以顺序地执行也可以不同的次序执行,只要能够实现本公开公开的技术方案所期望的结果,本文在此不进行限制。

[0158] 上述具体实施方式,并不构成对本公开保护范围的限制。本领域技术人员应该明白的是,根据设计要求和因素,可以进行各种修改、组合、子组合和替代。任何在本公开的精神和原则之内所作的修改、等同替换和改进等,均应包含在本公开保护范围之内。

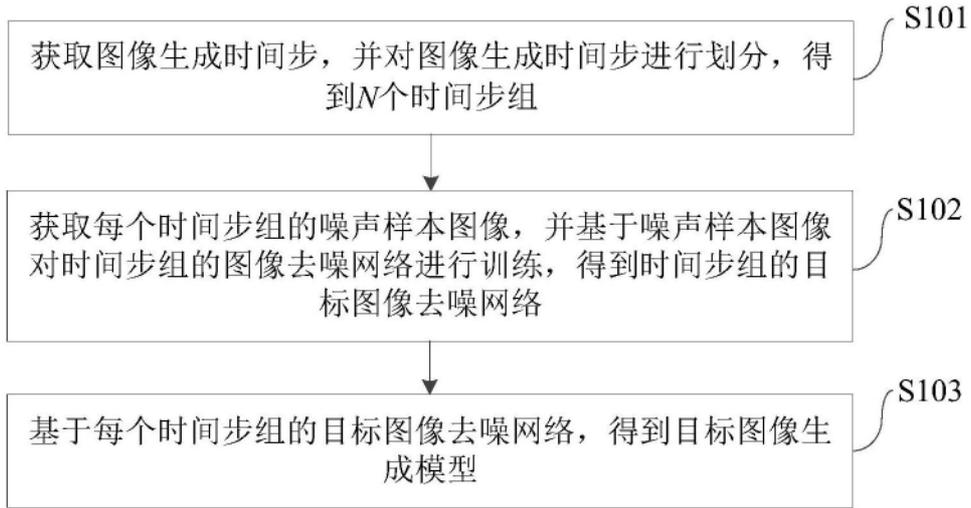


图1

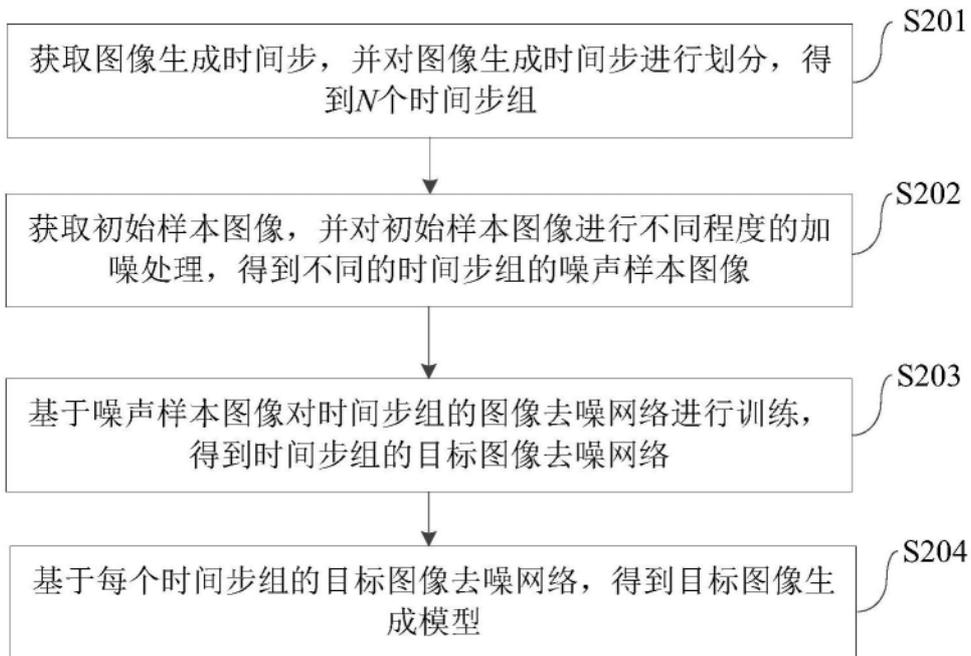


图2

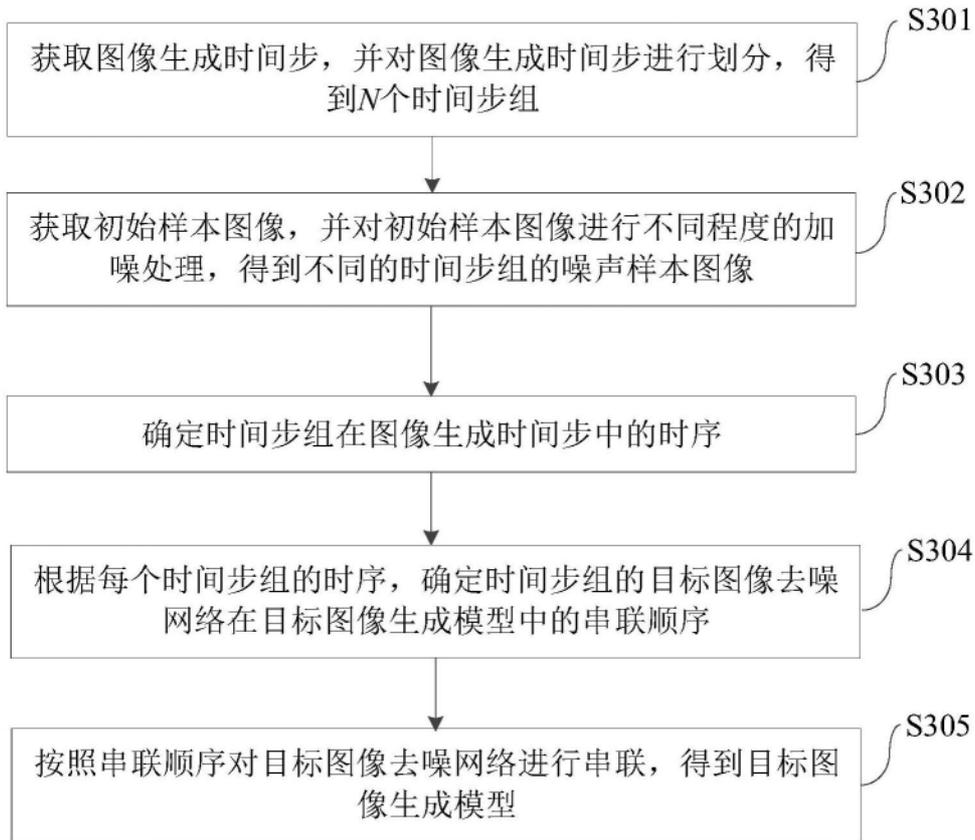


图3

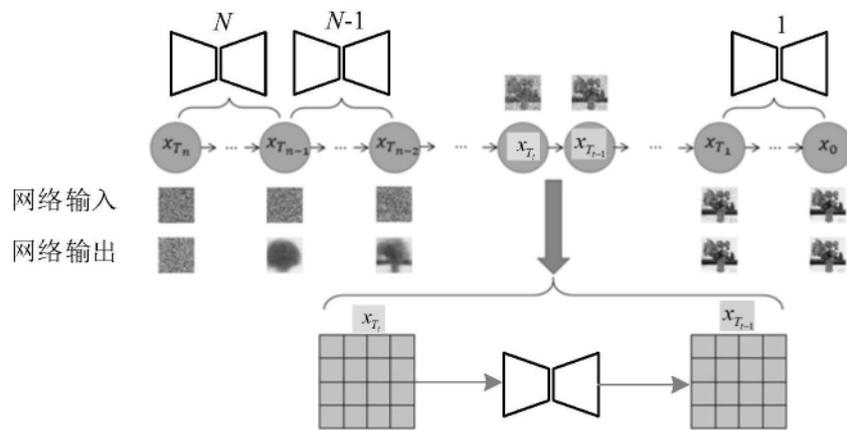


图4

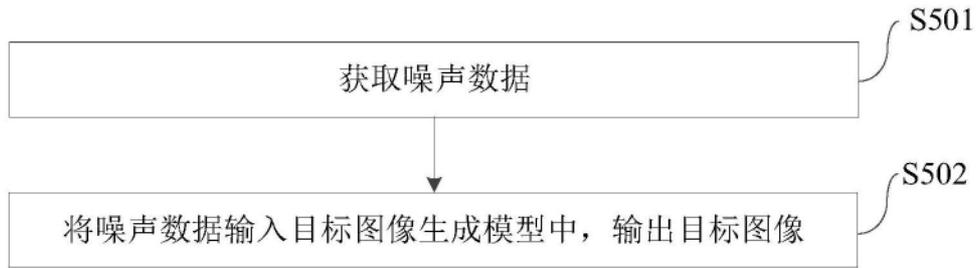


图5

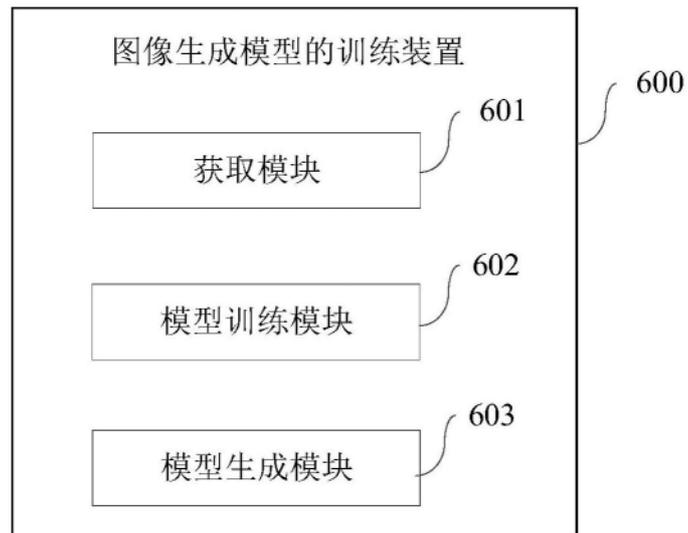


图6

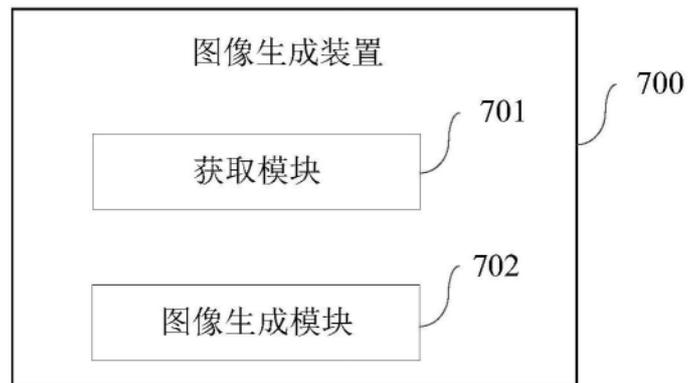


图7

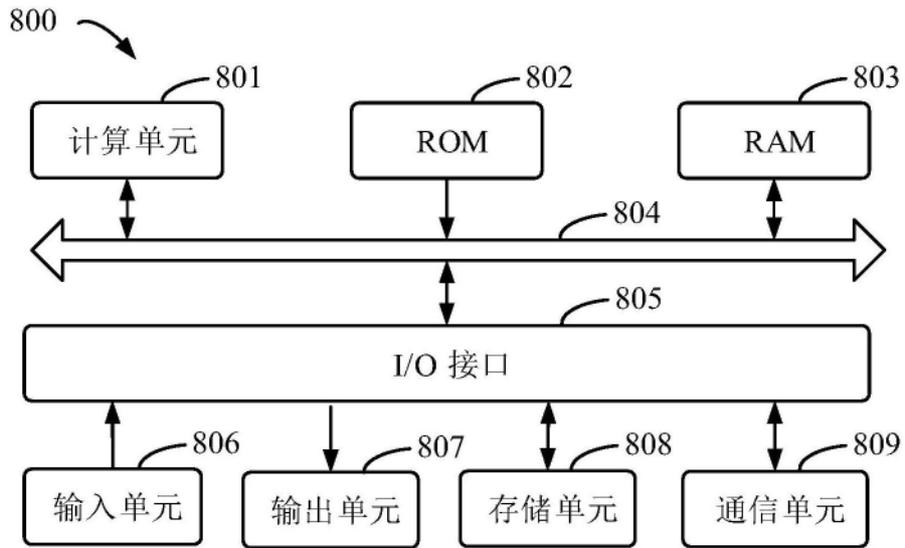


图8