(19)国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 117061766 A
(43)申请公布日 2023.11.14

(21)申请号 202311262525.8

(22)申请日 2019.09.16

(30)优先权数据
62/731,672 2018.09.14 US
16/254,475 2019.01.22 US

(62)分案原申请数据
201910868533.4 2019.09.16

(71)申请人 迪斯尼企业公司
地址 美国加利福尼亚州

(72)发明人 S·M·曼特 C·斯科尔斯 J·韩 S·D·伦巴多

(74)专利代理机构 北京纪凯知识产权代理有限公司 11245
专利代理师 李英

(51)Int.Cl.
H04N 19/44(2014.01)
H04N 19/124(2014.01)
H04N 19/91(2014.01)
G06T 9/00(2006.01)
G06N 3/08(2023.01)
G06N 3/0464(2023.01)

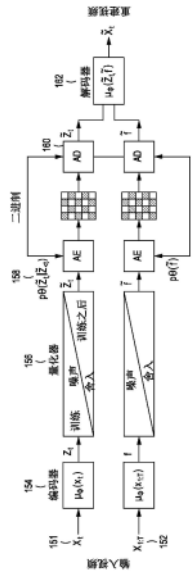
权利要求书4页 说明书17页 附图23页

(54)发明名称

基于机器学习的视频压缩

(57)摘要

本申请公开基于机器学习的视频压缩，并且公开用于压缩目标内容的系统和方法。在一个实施方式中，系统可以包括非瞬态电子存储设备和一个或多个物理计算机处理器。一个或多个物理计算机处理器可以由机器可读指令配置以获得包括一个或多个帧的目标内容，其中给定帧包括一个或多个特征。一个或多个物理计算机处理器可以由机器可读指令配置以获得条件网络。一个或多个物理计算机处理器可以由机器可读指令配置，以通过将条件网络应用于目标内容来生成解码的目标内容。



CN 117061766 A

1.一种被配置用于压缩目标内容的系统,所述系统包括:

非瞬态电子存储设备;以及

一个或多个物理计算机处理器,其由机器可读指令配置为:

从所述非瞬态电子存储设备获得包括一个或多个帧的所述目标内容,其中给定帧包括一个或多个特征;

从所述非瞬态电子存储设备获得条件网络,所述条件网络已通过使用训练内容训练初始网络来训练,其中所述条件网络包括一个或多个编码器、一个或多个量化器以及一个或多个解码器,并且其中所述训练内容包括一个或多个训练帧,并且其中给定训练帧包括一个或多个训练特征;

利用所述一个或多个物理计算机处理器,使用所述条件网络来编码所述目标内容,以生成一个或多个局部变量和一个或多个全局变量;

利用所述一个或多个物理计算机处理器,通过将所述条件网络应用于所述目标内容来生成所述目标内容的潜在空间,所述潜在空间包括所述一个或多个局部变量和所述一个或多个全局变量;

利用所述一个或多个物理计算机处理器,使用所述条件网络生成对应于所述潜在空间的多个分布;以及

利用所述一个或多个物理计算机处理器,使用所述条件网络基于所述多个分布量化所述潜在空间。

2.根据权利要求1所述的计算机实现的方法,其中所述一个或多个局部变量基于所述给定帧中的所述一个或多个特征,并且其中所述一个或多个全局变量基于所述目标内容的多个帧共有的一个或多个特征;以及

其中所述多个分布指示所述一个或多个局部变量和所述一个或多个全局变量的值的似然。

3.根据权利要求1所述的系统,其中编码所述目标内容包括:

利用所述一个或多个物理计算机处理器,将多个卷积层应用于所述目标内容以生成卷积的目标内容;

利用所述一个或多个物理计算机处理器,将全局模型应用于所述卷积的目标内容以生成所述一个或多个全局变量;以及

利用所述一个或多个物理计算机处理器,将多层感知器模型应用于所述卷积的目标内容以生成所述一个或多个局部变量。

4.根据权利要求3所述的系统,其中所述全局模型包括长短期记忆模型和卡尔曼滤波器中的一个或多个。

5.根据权利要求1所述的系统,其中应用所述条件网络还包括:

利用所述一个或多个物理计算机处理器,通过使用所述多个分布对所量化的潜在空间进行编码以生成经编码的量化的潜在空间;以及

利用所述一个或多个物理计算机处理器对所述经编码的量化的潜在空间进行解码。

6.根据权利要求5所述的系统,其中对所述经编码的量化的潜在空间进行解码包括:

利用所述一个或多个物理计算机处理器,对所述经编码的量化的潜在空间进行熵解码;

利用所述一个或多个物理计算机处理器,将经熵解码的潜在空间与多层感知器模型组合;以及

利用所述一个或多个物理计算机处理器,将多个反卷积应用于所述经熵解码的潜在空间与所述多层感知器模型的组合。

7.根据权利要求1所述的系统,其中对应于所述潜在空间的所述多个分布包括噪声并且以所述一个或多个全局变量和所述一个或多个局部变量的平均值为中心。

8.一种用于训练初始网络以同时学习如何使用训练内容来改进潜在空间以及如何使用所述训练内容来改进所述潜在空间的多个分布的计算机实现的方法,所述方法在包括非瞬态电子存储设备和一个或多个物理计算机处理器的计算机系统中实现,所述方法包括:

从所述非瞬态电子存储设备获得包括一个或多个训练帧的训练内容,其中给定训练帧包括一个或多个训练特征;

从所述非瞬态电子存储设备获得所述初始网络,所述初始网络包括一个或多个编码器、一个或多个量化器以及一个或多个解码器;

利用所述一个或多个物理计算机处理器,通过使用所述训练内容训练所述初始网络来生成条件网络,所述条件网络包括所述一个或多个编码器、所述一个或多个量化器以及所述一个或多个解码器;

在所述非瞬态电子存储设备中存储所述条件网络;

从所述非瞬态电子存储设备获得包括一个或多个帧的所述目标内容,其中给定帧包括一个或多个特征;

利用所述一个或多个物理计算机处理器使用所述条件网络来编码所述目标内容,以生成一个或多个局部变量和一个或多个全局变量;

利用所述一个或多个物理计算机处理器,使用所述条件网络生成所述潜在空间,所述潜在空间包括所述一个或多个局部变量和所述一个或多个全局变量;

利用所述一个或多个物理计算机处理器,使用所述条件网络生成对应于所述潜在空间的所述多个分布;以及

利用所述一个或多个物理计算机处理器,使用所述条件网络基于所述多个分布量化所述潜在空间。

9.根据权利要求8所述的计算机实现的方法,其中所述一个或多个局部变量基于所述给定帧中的所述一个或多个特征,并且其中所述一个或多个全局变量基于所述目标内容的多个帧共有的一个或多个特征;

其中所述多个分布指示所述一个或多个局部变量和所述一个或多个全局变量的值的多个似然。

10.根据权利要求8所述的计算机实现的方法,其中编码所述目标内容包括:

利用所述一个或多个物理计算机处理器,将多个卷积层应用于所述目标内容以生成卷积的目标内容;

利用所述一个或多个物理计算机处理器,将长短期记忆模型应用于所述卷积的目标内容以生成所述一个或多个全局变量;以及

利用所述一个或多个物理计算机处理器,将多层感知器模型应用于所述卷积的目标内容以生成所述一个或多个局部变量。

11.根据权利要求8所述的计算机实现的方法,还包括:

利用所述一个或多个物理计算机处理器,编码量化的潜在空间;以及

利用所述一个或多个物理计算机处理器,解码经编码的量化的潜在空间。

12.根据权利要求11所述的计算机实现的方法,其中解码所述经编码的量化的潜在空间包括:

利用所述一个或多个物理计算机处理器,对所述经编码的量化的潜在空间进行熵解码;

利用所述一个或多个物理计算机处理器,将经熵解码的潜在空间与多层感知器模型组合;以及

利用所述一个或多个物理计算机处理器,将多个反卷积应用于所述经熵解码的潜在空间与所述多层感知器模型的组合。

13.根据权利要求8所述的计算机实现的方法,其中对应于所述潜在空间的所述多个分布包括噪声并且以所述一个或多个全局变量和所述一个或多个局部变量的平均值为中心。

14.一种用于压缩目标内容的计算机实现的方法,所述方法在包括非瞬态电子存储设备和一个或多个物理计算机处理器的计算机系统中实现,所述方法包括:

从所述非瞬态电子存储设备获得包括一个或多个帧的所述目标内容,其中给定帧包括一个或多个特征;

利用所述一个或多个物理计算机处理器编码所述目标内容,以生成一个或多个局部变量和一个或多个全局变量;

利用所述一个或多个物理计算机处理器生成潜在空间,所述潜在空间包括所述一个或多个局部变量和所述一个或多个全局变量,其中所述一个或多个局部变量基于所述给定帧中的所述一个或多个特征,并且其中所述一个或多个全局变量基于所述目标内容的多个帧共有的一个或多个特征;

利用所述一个或多个物理计算机处理器,生成对应于所述潜在空间的多个分布;以及

利用所述一个或多个物理计算机处理器,基于所述多个分布量化所述潜在空间。

15.根据权利要求14所述的计算机实现的方法,其中所述多个分布指示所述一个或多个局部变量和所述一个或多个全局变量的值的似然。

16.根据权利要求14所述的计算机实现的方法,其中编码所述目标内容包括:

利用所述一个或多个物理计算机处理器,将多个卷积层应用于所述目标内容以生成卷积的目标内容;

利用所述一个或多个物理计算机处理器,将长短期记忆模型应用于所述卷积的目标内容以生成所述一个或多个全局变量;以及

利用所述一个或多个物理计算机处理器,将多层感知器模型应用于所述卷积的目标内容以生成所述一个或多个局部变量。

17.根据权利要求14所述的计算机实现的方法,还包括:

利用所述一个或多个物理计算机处理器,编码量化的潜在空间;以及

利用所述一个或多个物理计算机处理器,解码经编码的量化的潜在空间。

18.根据权利要求17所述的计算机实现的方法,其中解码所述经编码的量化的潜在空间包括:

利用所述一个或多个物理计算机处理器,对所述经编码的量化的潜在空间进行熵解码;

利用所述一个或多个物理计算机处理器,将经熵解码的潜在空间与多层感知器模型组合;以及

利用所述一个或多个物理计算机处理器,将多个反卷积应用于所述经熵解码的潜在空间与所述多层感知器模型的组合。

19.根据权利要求14所述的计算机实现的方法,其中对应于所述潜在空间的所述多个分布包括噪声并且以所述一个或多个全局变量和所述一个或多个局部变量的平均值为中心。

20.根据权利要求14所述的计算机实现的方法,其中所述潜在空间包括对应于所述一个或多个全局变量的全局密度模型和对应于所述一个或多个局部变量的局部密度模型。

基于机器学习的视频压缩

[0001] 本申请是2019年9月16日提交的名称为“基于机器学习的视频压缩”的中国专利申请201910868533.4的分案申请。

[0002] 相关申请的交叉引用

[0003] 本申请要求于2018年9月14日提交的美国临时专利申请No.62/731,672的优先权，该专利申请以引用方式并入本文。

技术领域

[0004] 本公开总体上涉及视频压缩。

发明内容

[0005] 本公开的实施方式针对用于压缩视频的系统和方法。

[0006] 在一个实施方式中，系统可以经配置用于压缩内容。该系统可以包括非瞬态电子存储(storage)/存储设备(storage)和由机器可读指令配置为执行若干操作的一个或多个物理计算机处理器。一个操作可以是从非瞬态电子存储设备获得包括一个或多个帧的目标内容。给定帧可以包括一个或多个特征。另一个操作可以是从非瞬态电子存储设备获得条件网络。可以通过使用训练内容训练初始网络来训练条件网络。条件网络可以包括一个或多个编码器、一个或多个量化器以及一个或多个解码器。训练内容可以包括一个或多个训练帧。给定的训练帧可以包括一个或多个训练特征。另一个这样的操作可以是利用一个或多个物理计算机处理器，通过将条件网络应用于目标内容来生成解码的目标内容。条件网络可以生成目标内容的潜在空间(latent space)。目标内容可以包括一个或多个局部变量和一个或多个全局变量。

[0007] 在实施方式中，应用条件网络可以包括利用一个或多个物理计算机处理器，通过使用条件网络编码目标内容以生成一个或多个局部变量和一个或多个全局变量。应用条件网络可以包括利用一个或多个物理计算机处理器，通过使用条件网络生成潜在空间。潜在空间可以包括一个或多个局部变量和一个或多个全局变量。一个或多个局部变量可以基于给定帧中的一个或多个特征。一个或多个全局变量可以基于目标内容的多个帧共有的一个或多个特征。应用条件网络可以包括利用一个或多个物理计算机处理器，通过使用条件网络生成对应于潜在空间的多个分布。多个分布指示一个或多个局部变量和一个或多个全局变量的值的似然(likelihood)。应用条件网络可以包括利用一个或多个物理计算机处理器量化潜在空间。

[0008] 在实施方式中，编码目标内容可以包括利用一个或多个物理计算机处理器，将多个卷积层应用于目标内容。编码目标内容可以包括利用一个或多个物理计算机处理器，将全局模型应用于卷积的目标内容以生成一个或多个全局变量。编码目标内容可以包括利用一个或多个物理计算机处理器，将多层感知器模型应用于卷积的目标内容以生成一个或多个局部变量。

[0009] 在实施方式中，全局模型包括长短期记忆模型和卡尔曼滤波器中的一个或多个。

[0010] 在实施方式中,应用条件网络还可以包括利用一个或多个物理计算机处理器,通过使用多个分布对量化的潜在空间进行编码。应用条件网络还可以包括利用一个或多个物理计算机处理器对经编码的潜在空间进行解码。

[0011] 在实施方式中,解码量化的潜在空间可以包括利用一个或多个物理计算机处理器,对经编码的潜在空间进行熵解码。解码量化的潜在空间可以包括利用一个或多个物理计算机处理器,将经熵解码的潜在空间与多层感知器模型组合。解码量化的潜在空间可以包括利用一个或多个物理计算机处理器,将多个反卷积应用于经熵解码的潜在空间与多层感知器模型的组合。

[0012] 在实施方式中,对应于潜在空间的多个分布以一个或多个全局变量和一个或多个局部变量的平均值为中心,并且包括噪声。

[0013] 在另一个实施方式中,可以在计算机系统中实现计算机实现的方法,计算机实现的方法用于训练初始网络以同时学习如何使用训练内容来改进潜在空间以及如何使用训练内容来改进潜在空间的多个分布。计算机系统可以包括非瞬态电子存储设备和一个或多个物理计算机处理器。计算机实现的方法可以包括从非瞬态电子存储设备获得包括一个或多个帧的训练内容。计算机实现的方法还可以包括从非瞬态电子存储设备获得初始网络。初始网络可以包括一个或多个编码器、一个或多个量化器以及一个或多个解码器。计算机实现的方法可以包括利用一个或多个物理计算机处理器,通过使用训练内容训练初始网络来生成条件网络。条件网络可以包括一个或多个编码器、一个或多个量化器以及一个或多个解码器。计算机实现的方法可以包括在非瞬态电子存储设备中存储条件网络。

[0014] 在实施方式中,计算机实现的方法还可以包括从非瞬态电子存储设备获得包括一个或多个帧的目标内容。给定帧可以包括一个或多个特征。计算机实现的方法可以包括利用一个或多个物理计算机处理器编码目标内容,以使用条件网络生成一个或多个局部变量和一个或多个全局变量。计算机实现的方法可以包括利用一个或多个物理计算机处理器,通过使用条件网络生成潜在空间。潜在空间可以包括一个或多个局部变量和一个或多个全局变量。一个或多个局部变量基于给定帧中的一个或多个特征。一个或多个全局变量基于目标内容的多个帧共有的一个或多个特征。计算机实现的方法可以包括利用一个或多个物理计算机处理器,通过使用条件网络生成对应于潜在空间的多个分布。多个分布可以指示一个或多个局部变量和一个或多个全局变量的值的多个似然。计算机实现的方法可以包括利用一个或多个物理计算机处理器,通过使用条件网络基于多个分布量化潜在空间。

[0015] 在实施方式中,编码目标内容可以包括利用一个或多个物理计算机处理器,将多个卷积层应用于目标内容。编码目标内容可以包括利用一个或多个物理计算机处理器,将长短期记忆模型应用于卷积的目标内容以生成一个或多个全局变量。编码目标内容可以包括利用一个或多个物理计算机处理器,将多层感知器模型应用于卷积的目标内容以生成一个或多个局部变量。

[0016] 在实施方式中,计算机实现的方法还可以包括利用一个或多个物理计算机处理器,编码量化的潜在空间。计算机实现的方法可以包括利用一个或多个物理计算机处理器,解码经编码的潜在空间。

[0017] 在实施方式中,解码量化的潜在空间可以包括利用一个或多个物理计算机处理器,对经编码的潜在空间进行熵解码。解码量化的潜在空间可以包括利用一个或多个物理

计算机处理器，将经熵解码的潜在空间与多层感知器模型组合。解码量化的潜在空间可以包括利用一个或多个物理计算机处理器，将多个反卷积应用于经熵解码的潜在空间与多层感知器模型的组合。

[0018] 在实施方式中，对应于潜在空间的多个分布以一个或多个全局变量和一个或多个局部变量的平均值为中心，并且包括噪声。

[0019] 在一个实施方式中，用于压缩目标内容的计算机实现的方法可以在包括非瞬态电子存储设备和一个或多个物理计算机处理器的计算机系统中实现。计算机实现的方法可以包括从非瞬态电子存储设备获得包括一个或多个帧的目标内容。给定帧包括一个或多个特征。计算机实现的方法可以包括利用一个或多个物理计算机处理器编码目标内容，以生成一个或多个局部变量和一个或多个全局变量。计算机实现的方法可以包括利用一个或多个物理计算机处理器生成潜在空间，该潜在空间包括一个或多个局部变量和一个或多个全局变量。一个或多个局部变量基于给定帧中的一个或多个特征。一个或多个全局变量基于目标内容的多个帧共有的一个或多个特征。

[0020] 在实施方式中，计算机实现的方法还可以包括利用一个或多个物理计算机处理器，生成对应于潜在空间的多个分布。多个分布可以指示一个或多个局部变量和一个或多个全局变量的值的似然。计算机实现的方法可以包括利用一个或多个物理计算机处理器，基于多个分布量化潜在空间。

[0021] 在实施方式中，编码目标内容可以包括利用一个或多个物理计算机处理器，将多个卷积层应用于目标内容。编码目标内容可以包括利用一个或多个物理计算机处理器，将长短期记忆模型应用于卷积的目标内容以生成一个或多个全局变量。编码目标内容可以包括利用一个或多个物理计算机处理器，将多层感知器模型应用于卷积的目标内容以生成一个或多个局部变量。

[0022] 在实施方式中，计算机实现的方法还可以包括利用一个或多个物理计算机处理器，编码量化的潜在空间。计算机实现的方法可以包括利用一个或多个物理计算机处理器，解码经编码的潜在空间。

[0023] 在实施方式中，解码量化的潜在空间可以包括利用一个或多个物理计算机处理器，对经编码的潜在空间进行熵解码。解码量化的潜在空间可以包括利用一个或多个物理计算机处理器，将经熵解码的潜在空间与多层感知器模型组合。解码量化的潜在空间可以包括利用一个或多个物理计算机处理器，将多个反卷积应用于经熵解码的潜在空间与多层感知器模型的组合。

[0024] 在实施方式中，对应于潜在空间的多个分布以一个或多个全局变量和一个或多个局部变量的平均值为中心，并且随机噪声被添加到多个分布。

[0025] 在实施方式中，潜在空间包括对应于一个或多个全局变量的全局密度模型和对应于一个或多个局部变量的局部密度模型。

[0026] 在实施方式中，全局密度模型由下式定义：

[0027] $$p_\theta(f) = \prod_{i}^{\dim(f)} p_\theta(f^i) * \mathcal{U}\left(-\frac{1}{2}, \frac{1}{2}\right)$$

[0028] 其中p_θ表示密度模型，f表示一个或多个全局变量，i表示与一个或多个全局变量

的维度对应的维度索引,并且$\mathcal{U}\left(-\frac{1}{2},\frac{1}{2}\right)$表示随机噪声,并且其中局部密度模型由下式定义:

[0029] $$p_\theta(z_{1:T}) = \prod_{i}^{T} \prod_{i}^{\dim(z)} p_\theta(z_t^i|c_t) * \mathcal{U}\left(-\frac{1}{2},\frac{1}{2}\right)$$

[0030] 其中p_θ表示密度模型,z表示一个或多个局部变量,T表示对应于给定帧的时间,i表示与一个或多个全局变量的维度对应的维度索引,c_t表示对应于T的上下文参数,并且$\mathcal{U}\left(-\frac{1}{2},\frac{1}{2}\right)$表示随机噪声。

附图说明

[0031] 本专利或申请文件包含至少一幅彩色附图。具有(一幅或多幅)彩色附图的本专利或专利申请出版物的副本将在请求并支付必要费用后由事务所提供。

[0032] 当结合附图阅读下面描述的各种公开实施方式的详细描述时,将理解本公开的各方面。

[0033] 图1A示出根据各种实施方式的用于压缩内容的示例性系统。

[0034] 图1B示出根据本公开的各种实施方式的用于压缩内容的模型的示例性架构。

[0035] 图2是示出根据一个实施方式的用于压缩内容的示例性过程的操作流程图。

[0036] 图3A示出根据本公开的各种实施方式的使用当前公开技术的示例性压缩的目标内容。

[0037] 图3B示出使用H.265的示例性压缩的目标内容。

[0038] 图3C示出使用VP9的示例性压缩的目标内容。

[0039] 图4A示出根据本公开的各种实施方式的将当前公开技术与现有技术进行比较的示例性压缩的目标内容。

[0040] 图4B示出根据本公开的各种实施方式的将当前公开技术与现有技术进行比较的示例性压缩的目标内容。

[0041] 图5A示出根据本公开的各种实施方式的公开技术在视频数据集方面的示例性能。

[0042] 图5B示出根据本公开的各种实施方式的公开技术在视频数据集方面的示例性能。

[0043] 图6A示出根据本公开的各种实施方式的公开技术在视频数据集方面的示例性能。

[0044] 图6B示出根据本公开的各种实施方式的公开技术在视频数据集方面的示例性能。

[0045] 图7A示出根据本公开的各种实施方式的使用多个压缩模型的示例性率失真曲线。

[0046] 图7B示出根据本公开的各种实施方式的使用多个压缩模型的示例性率失真曲线。

[0047] 图7C示出根据本公开的各种实施方式的使用多个压缩模型的示例性率失真曲线。

[0048] 图8A示出根据本公开的各种实施方式的使用多个压缩模型的示例性率失真曲线。

[0049] 图8B示出根据本公开的各种实施方式的使用多个压缩模型的示例性率失真曲线。

[0050] 图8C示出根据本公开的各种实施方式的使用多个压缩模型的示例性率失真曲线。

[0051] 图9A示出根据本公开的各种实施方式的使用多个压缩模型的示例性信息平均比特。

[0052] 图9B示出根据本公开的各种实施方式的使用多个压缩模型的示例性信息平均比特。

[0053] 图9C示出根据本公开的各种实施方式的使用多个压缩模型的示例性信息平均比特。

[0054] 图10A示出根据本公开的各种实施方式的使用当前公开技术的示例性分布。

[0055] 图10B示出根据使用本公开的各种实施方式的使用当前公开技术的示例性分布。

[0056] 图10C示出根据使用本公开的各种实施方式的当前公开技术的示例性分布。

[0057] 图10D示出根据本公开的各种实施方式的使用当前公开技术的示例性分布。

[0058] 图11示出可以用于实现本公开的各种实施方式的特征的示例性计算部件。

[0059] 在下面的描述和示例中更详细地描述附图,这些附图仅为了说明的目的而提供,并且仅描述本公开的示例性实施方式。这些附图并不旨在穷举或将本公开限制于所公开的精确形式。还应理解,本公开可以通过修改或变更来实践,并且本公开可以仅由权利要求书及其等同物来限制。

具体实施方式

[0060] 本公开涉及用于基于机器学习的视频压缩的系统和方法。视频压缩可以包括无损压缩和有损压缩。例如,无损视频压缩可以与概率建模相关联。序列中后续视频帧的准确概率性知识可以最大限度地压缩原始视频中的信息内容。有损压缩可以包括丢弃不相关的图像信息以减小文件大小,同时感知的图像质量保持可接受。丢弃信息的过程可以过滤与感知的视频质量无关的信息。变分自编码器(VAE)框架的现有用途已经应用于单个图像压缩应用,并且使用机器学习(即、深度学习)的现有视频压缩框架仅关注帧插值。当前公开的技术使用媒体帧的时变概率分布来实现更高的压缩比。

[0061] 现有编解码器可以使用关键帧,这些关键帧与图像压缩一起存储并用于内插到中间帧。当前公开的技术可以更连贯地使用所有帧并决定应存储来自整个序列的哪些信息,而不是挑出某些帧作为关键帧。用于视频压缩的视频编解码器通常将视频分解成被编码为单个图像的一组关键帧以及使用插值的一组帧。相反,本公开应用深度学习(例如,神经网络)来编码、压缩和解码视频。

[0062] 在实施方式中,压缩编解码器可以包括编码器和解码器。编码器可以将媒体内容转换成比特串,并且解码器可以从比特中恢复原始媒体内容。编解码器可以经配置基于用户选择的图像质量将媒体转换成可能的最小比特数。为了最大限度地压缩源,编码器和解码器都可以在写入/写出和读入/读出压缩格式时使用预测模型。当前公开的技术可以使用深度生成模型,该模型能够从媒体源学习图像的顺序概率分布,并且可以在任何给定时间用作下一视频帧的预测模型。图像编解码器例如JPEG或GIF可以独立地编码每个图像,而视频编解码器例如MPEG-4part2、H.264或H.265可以编码整个视频帧序列。

[0063] 在实施方式中,预测模型可以包括用于执行视频压缩的VAE框架。VAE框架的编码器部分可以用于将媒体帧序列编码成潜在表示(例如,可以用于解码媒体内容的压缩数据)。潜在空间可以在压缩时间离散化,使得潜在变量可以转换成压缩格式,例如二进制。通过使用VAE的解码器部分,可以从压缩格式中恢复原始媒体序列,该解码器部分解码来自压缩格式的潜在变量并从潜在变量产生原始视频帧的近似。

[0064] 本文公开的其它实施方式涉及使用机器学习的媒体压缩以减小媒体文件大小。例如,可以将各个视频帧转换成可以滤除无关信息的表示,并且视频的时间方面可以被概率地建模以减小压缩文件长度。可以通过利用连续变分自编码器在大视频数据集上训练模型,生成最佳图像表示和时间相关的概率分布。如本文将描述的,VAE可以针对复杂概率分布使用无监督学习的形式。

[0065] 在详细描述该技术之前,描述可以实现当前公开技术的示例性环境可能是有用的。图1A示出一个这样的示例性环境100。

[0066] 环境100可以结合实现所公开的系统、方法和设备的实施方式来使用。作为示例,图1A的各种下述部件可以通过编码目标内容、生成潜在空间、生成一个或多个概率分布、量化潜在空间、编码量化的潜在空间和/或解码潜在空间,用于压缩目标内容。目标内容可以包括图像、视频、数据和/或其它内容。目标内容可以包括一个或多个特征和/或其它信息。潜在空间可以包括一个或多个局部变量和一个或多个全局变量。一个或多个局部变量可以基于目标内容的单个帧和/或基于帧内空间相关性。一个或多个全局变量可以基于目标内容的多个帧、整个目标内容和/或基于目标内容的多个帧之间的时间相关性。服务器系统106可以包括表示模型114、编码器116、量化器118和解码器120,如本文将描述的。表示模型114可以学习如何使用训练内容来改进潜在空间以及如何使用训练内容来改进潜在空间的多个分布。编码器116可以转换目标内容以生成潜在空间,并且可以编码与潜在空间对应的一个或多个概率分布。量化器116可以基于一个或多个概率分布来量化和/或舍入一个或多个潜在空间值。解码器120可以解码潜在空间,并且可以将潜在空间转换回目标内容的格式。

[0067] 可以训练和/或调节表示模型114以压缩目标内容。表示模型114可以包括神经网络、深层生成模型、长短期记忆(LSTM)模型、多层感知器(MLP)模型、卡尔曼滤波器模型、量化器、VAE及对应的解码器、熵编码器及对应的解码器、算术编码器及对应的解码器、机器学习模型等中的一个或多个。初始表示模型可以对训练内容使用无监督训练,以改进以下方法:将训练内容转换成包括一个或多个局部变量和/或一个或多个全局变量的潜在空间的方法,生成对应于目标内容的一个或多个帧的一个或多个概率分布,和/或使用一个或多个先前帧的一个或多个潜在变量和/或一个或多个概率分布来预测和/或生成一个或多个后续帧的一个或多个潜在变量和/或一个或多个概率分布。初始表示模型可以通过将解码的训练内容的一个或多个参数与原始训练内容的一个或多个参数进行比较并使用该比较作为反馈来训练,以调节和/或训练初始表示模型。可以使用从存储/存储设备110获得的训练数据在服务器系统106中训练初始模型。初始模型可以使用机器学习(例如,神经网络和/或其它机器学习算法)来训练模型114并改善模型114的输出。得到的条件模型可以存储在存储设备110中。

[0068] 条件模型能够将目标内容转换成一个或多个局部变量和一个或多个全局变量,并生成一个或多个概率分布以进行编码。在编码之后,条件模型可以对编码的目标内容进行转换并解码。应理解,可以使用其它训练来训练表示模型114。

[0069] 图1B是根据本公开各种实施方式的示例性条件表示模型。示例性表示模型可以包括一个或多个编码器154和158、量化器156以及解码器160和162。输入151和152可以包括分别分成单个帧和包括多个帧的段的目标内容。编码器154可以基于训练内容从输入151提

取/生成一个或多个局部变量z_t并从输入152提取/生成一个或多个全局变量f。如本文所述,训练可以改善来自目标内容的一个或多个潜在变量的生成。编码器154可以生成对应于一个或多个局部变量和一个或多个全局变量的一个或多个概率分布p_θ。可以基于均匀分布生成一个或多个概率分布。应理解,可以使用其它分布来生成一个或多个概率分布(例如,高斯分布)。在一些实施方式中,一个或多个概率分布可以以一个或多个潜在变量的平均值为中心。可以基于编码器154生成和/或预测一个或多个变量的平均值。应理解,可以选择预测其它统计值和/或可以基于机器学习预测其它值。如本文所述,训练可以改善一个或多个概率分布的生成。

[0070] 量化器156可以将噪声添加或注入到一个或多个潜在变量和/或一个或多个概率分布,和/或量化概率分布、一个或多个局部变量和/或一个或多个全局变量。编码器158可以将一个或多个概率分布和/或一个或多个变量熵编码、算术编码和/或以其它方式无损编码成二进制。这个过程可以减少一个或多个冗余的潜在变量。解码器160可以从二进制中熵解码一个或多个编码的概率分布和/或一个或多个编码的变量。解码器162可以解码和/或转换一个或多个熵解码的概率分布和/或一个或多个熵解码的潜在变量,以生成解码的目标内容。

[0071] 编码器154和158以及解码器160和162可以使用概率模型,如本文所述,该概率模型在大视频数据集的训练期间与编码器和解码器一起被学习。概率模型可以称为先验概率或帧分布。先验概率可以是给定先前视频帧的下一视频帧的概率分布。概率分布可以预测似乎合理的未来视频帧的分布。在实施方式中,模型114可以使用从第一帧学习的信息来预测,或生成一个或多个后续帧的一个或多个潜在变量以及一个或多个概率分布。

[0072] 在一些实施方式中,先验模型可以依赖于一个或多个先前帧(例如,深度卡尔曼滤波器)。如本文所述,先验模型可以将变量分解成全局变量和局部变量。几个帧共有的信息可以存储为全局变量,每个帧的新信息可以存储在局部变量中。可以使用完全连接的神经网络来确定信息内容是应存储在局部变量中还是全局变量中,以基于一个或多个先前帧的一个或多个潜在变量预测一个或多个后续帧。

[0073] 在一些实施方式中,先验模型可以包括长短期记忆(LSTM)模型,以说明多个先前帧,而不是单个先前视频帧。该模型可以具有比卡尔曼滤波器模型更长的记忆,并且考虑更长的视频帧序列以预测下一帧。在一些实施方式中,LSTM模型可以是双向的,其说明给定帧之前和之后的视频帧序列,以便预测给定帧的概率分布。具有双向LSTM模型的模型可以具有用于预测的最大信息量(例如,过去帧和未来帧)。在一些实施方式中,双向LSTM模型可能需要将附加的辅助信息(side information)存储成压缩格式。解码器可以使用附加的辅助信息,基于概率分布顺序地解码帧(这取决于尚未解码的未来帧)。双向LSTM的隐藏状态可以包括关于未来帧的这种信息。在该变型中,LSTM的隐藏状态可以被离散化并存储在压缩视频文件中。

[0074] 在实施方式中,LSTM模型的隐藏状态可以用于预测下一帧的潜在变量。LSTM的隐藏状态可以用于预测下一帧的潜在变量。在一些实施方式中,隐藏状态可以编码下一帧的潜在变量。在另一示例中,LSTM的隐藏状态可以与前一帧的潜在变量组合,以生成下一视频帧。LSTM的隐藏状态可以用于将前一帧的潜在变量转换或生成为下一帧的潜在变量。

[0075] 在一些实施方式中,先验模型可以包括使用运动信息来预测宏观运动和/或其它

运动的运动表示模型。在实施方式中,运动信息可以被非线性编码在视频的潜在表示的结构中,以改善压缩。可以基于神经网络的大小,针对不同的分辨率媒体缩放先验模型。

[0076] 返回参考图1A,如上所述,编码器116可以转换内容、编码内容和/或生成多个分布。如上所述,量化器118可以注入噪声和/或量化目标内容。如上所述,解码器120可以将潜在空间解码并转换成目标内容的格式。

[0077] 电子设备102可以包括各种电子计算设备,例如智能手机、平板电脑、笔记本电脑、计算机、可穿戴设备、电视、虚拟现实设备、增强现实设备、显示器、连接的家庭设备、物联网(IOT)设备、智能扬声器和/或其它设备。电子设备102可以向用户呈现内容和/或接收将内容发送给另一用户的请求。在一些实施方式中,电子设备102可以将表示模型114、编码器116、量化器118和/或解码器120应用于目标内容。在实施方式中,电子设备102可以存储表示模型114、编码器116、量化器118和解码器120。

[0078] 如图1A所示,环境100可以包括电子设备102和服务器系统106中的一个或多个。电子设备102可以经由通信介质104耦合到服务器系统106。如本文将详细描述的,电子设备102和/或服务器系统106可以经由通信介质104交换通信信号,包括内容、一个或多个特征、一个或多个潜在空间、一个或多个局部变量、一个或多个全局变量、一个或多个模型、元数据、用户输入、辅助信息和/或其它信息。

[0079] 在各种实施方式中,通信介质104可以基于一个或多个无线通信协议,诸如Wi-Fi(无线网络)、Bluetooth®、ZigBee、802.11协议、红外(IR)、射频(RF)、2G、3G、4G、5G等,和/或有线协议和介质。在某些情况下,通信介质104可以被实现为单个介质。

[0080] 如上所述,通信介质104可以用于将电子设备102和/或服务器系统106彼此连接或通信耦合或连接或通信耦合到网络,并且通信介质104可以以多种形式实现。例如,通信介质104可以包括互联网连接,诸如局域网(LAN)、广域网(WAN)、光纤网络、互联网电源线(internet over power lines)、硬线连接(例如,总线)等,或者任何其它类型的网络连接。可以使用路由器、电缆、调制解调器、交换机、光纤、电线、无线电(例如,微波/RF链路)等的任意组合来实现通信介质104。在阅读本公开之后,应理解,可以使用其它方式来实现通信介质104以用于通信目的。

[0081] 同样,应理解,除了环境100的其它元件之外,类似的通信介质可以用于将服务器108、存储设备110、处理器112、表示模型114、编码器116、量化器118和/或解码器120彼此连接或通信耦合。在示例性实施方式中,通信介质104可以是或包括用于电子设备102和/或服务器系统106的有线或无线广域网(例如,蜂窝、光纤和/或电路交换连接等),其可以在地理上相对不同;并且在一些情况下,通信介质104的各方面可以涉及有线或无线局域网(例如,Wi-Fi(无线网络)、Bluetooth(蓝牙)、未许可的无线连接、USB(通用串行总线)、HDMI(高清晰度多媒体接口)、标准AV等),其可以用于通信地耦合环境100的可以在地理上相对靠近的各方面。

[0082] 服务器系统106可以向/从电子设备102提供、接收、收集或监控信息,诸如内容、一个或多个特征、一个或多个潜在空间、一个或多个局部变量、一个或多个全局变量、一个或多个模型、元数据、用户输入、安全和加密信息、辅助信息等。服务器系统106可以经配置经由通信介质104接收或发送这样的信息。该信息可以存储在存储设备110中,并且可以使用处理器112来处理。例如,处理器112可以包括能够对服务器系统106已从电子设备102收集、

接收等的信息执行分析的分析引擎。处理器112可以包括表示模型114、编码器116、量化器118和解码器120，其能够接收内容、编码内容、量化内容、解码内容、分析内容以及以其它方式处理服务器系统106基于请求已从电子设备102收集、接收等的内容。在实施方式中，服务器108、存储设备110和处理器112可以实现为分布式计算网络、关系数据库等。

[0083] 服务器108可以包括例如互联网服务器、路由器、台式或膝上型计算机、智能手机、平板电脑、处理器、部件等，并且可以以各种形式实现，包括例如集成电路或其集合、印刷电路板或其集合、或者分立的外壳/封装/机架或其多个。服务器108可以更新存储在电子设备102上的信息。服务器108可以实时或偶尔地向/从电子设备102发送/接收信息。进一步地，服务器108可以实现用于电子设备102的云计算能力。在研究本公开之后，本领域技术人员将理解，环境100可以包括多个电子设备102、通信介质104、服务器系统106、服务器108、存储设备110、处理器112、表示模型114、编码器116、量化器118和/或解码器部件120。

[0084] 图2是示出根据一个实施方式的用于压缩目标内容的示例性过程的操作流程图。本文描述的各种方法的操作不一定限于图中描述或示出的顺序，并且在研究本公开之后，应理解，本文描述的操作顺序的变化在本公开的精神和范围内。

[0085] 在一些情况下，可以通过系统100的部件、元件、设备、部件和电路中的一个或多个来执行流程图的操作和子操作。这可以包括以下中的一个或多个：本文描述并至少参考图1A、图1B和图11引用的服务器系统106；服务器108；处理器112；存储设备110；和/或计算部件1100以及其中描述和/或参考其引用的子部件、元件、设备、部件和电路。在这种情况下，流程图的描述可以指代对应的部件、元件等，但不管是否进行明确的引用，在研究本公开之后，应理解，可以使用对应的部件、元件等。进一步地，应理解，此类引用不一定将所描述的方法限制于所提及的特定部件、元件等。因此，应理解，上面结合(子)部件、元件、设备、电路等描述的方面和特征(包括其变型)可以应用于结合流程图描述的各种操作而不脱离本公开的范围。

[0086] 在操作202处，可以获得训练内容。训练内容可以包括图像、视频和/或其它媒体内容。训练内容可以包括一个或多个特征。一个或多个特征可以用于生成包括一个或多个局部变量和/或一个或多个全局变量的潜在空间。

[0087] 在操作204处，可以获得初始网络。如上所述，初始网络可以包括一个或多个编码器、量化器和/或解码器。

[0088] 在操作206处，可以通过使用训练内容训练初始网络来生成条件网络。如上所述，可以训练条件网络以改进用于生成一个或多个局部变量和/或一个或多个全局变量的编码过程。如上所述，可以训练条件网络以改进用于生成多个分布的编码过程。在实施方式中，条件网络可以使用来自一个或多个先前帧的一个或多个潜在变量和/或一个或多个概率分布来预测后续帧。

[0089] 在一个示例中，可以通过最小化证据下限来同时训练编码器和解码器。可以通过在下限中重新调节单个项来修改优化，以实现潜在表示的不同概率行为。该修改可以用于调整视频质量与压缩视频文件长度之间的折衷。例如，β编码器损耗可以由下式定义：

$$-\mathbb{E}_{\tilde{f},\tilde{z}_{1:T}\sim q}\left[logp_\theta(x_{1:T}|\tilde{f},\tilde{z}_{1:T})\right]-\beta\mathbb{E}_{\tilde{f},\tilde{z}_{1:T}\sim q}\left[logp_\theta(\tilde{f},\tilde{z}_{1:T})\right]$$

[0090]
$$=\mathbb{E}_{\tilde{f},\tilde{z}_{1:T}\sim q}\sum_{t=1}^{T}\left|\left|\tilde{x}_t-x_t\right|\right|_1+\beta H\left[q_\phi(\tilde{z}_{1:T},f|x_{1:T}),p_\theta(\tilde{f},\tilde{z}_{1:T})\right]$$

[0091] 其中重构帧$\tilde{x}_t=\mu_\theta(\mu_\phi(x_{1:T}))$，$\mathbb{E}$可以表示预期值(例如,预期值的加权平均值),并且其余变量可以表示与本文所述相同的功能。在一些实施方式中,具有βH的第二项可以在使用先验分布$p(f,z_{1:T})$对潜在空间进行熵编码时对应于预期代码长度。当代码的经验分布与先验模型(例如,$p(f,z_{1:T})=q(f,z_{1:T}|x_{1:T})$)匹配时,可以最小化第二项。在实施方式中,一个或多个全局变量的交叉熵可以由下式定义:

$$H\left[q_\phi(f|x_{1:T}),p_\theta(f)\right]$$

[0092]
$$=H\left[q_\phi(f|x_{1:T}),p_\theta(f^i)\right]=-\mathbb{E}_{f\sim q}\sum_{i=1}^{N}\log_2 p_\theta(f^i)+\beta H$$

[0093] 其中变量可以具有与上述相同的表示。一个或多个局部变量的交叉熵可以由下式定义:

$$H\left[q_\phi(z_{1:T}|x_{1:T}),p_\theta(z_{1:T})\right]$$

[0094]
$$=H\left[q_\phi(z_{1:T}|x_{1:T}),p_\theta(z_{1:T})\right]=-\mathbb{E}_{z_{1:T}\sim q}\sum_{t=1}^{T}\sum_{i=1}^{N}\log_2 p_\theta(z_t^i|c_t)$$

[0095] 其中变量可以具有与上述相同的表示,并且c_t可以表示上下文变量。

[0096] 在操作208处,可以将条件网络存储在例如非瞬态电子存储设备中。

[0097] 在操作210处,可以获得目标内容。目标内容可以包括图像、视频和/或其它媒体内容。目标内容可以包括一个或多个特征。一个或多个特征可以用于生成包括一个或多个局部变量和/或一个或多个全局变量的潜在空间。

[0098] 在操作212处,可以编码目标内容。编码目标内容可以生成一个或多个局部变量和一个或多个全局变量。在实施方式中,编码目标内容可以包括利用一个或多个物理计算机处理器将多个卷积层应用于目标内容。编码目标内容可以包括利用一个或多个物理计算机处理器将长短期记忆模型应用于卷积的目标内容,以生成一个或多个全局变量。编码目标内容还可以包括利用一个或多个物理计算机处理器将多层感知器模型应用于卷积的目标内容,以生成一个或多个局部变量。在一些实施方式中,可以用于生成一个或多个潜在变量的概率模型可以由下式定义:

[0099] $$p_\theta(x_{1:T},z_{1:T},f)=p_\theta(f)p_\theta(z_{1:T})\prod_{t=1}^{T}p_\theta(x_t|z_t,f)$$

[0100] 其中p_θ可以表示具有参数θ的概率,$x_{1:T}$可以表示时间T的目标内容的给定帧,$z_{1:T}$

可以表示时间T的一个或多个局部变量,并且f可以表示一个或多个全局变量。θ可以是生成模型的所有参数的简写符号。

[0101] 在操作214处,可以生成潜在空间。潜在空间可以包括一个或多个局部变量和/或一个或多个全局变量。一个或多个局部变量基于给定帧中的一个或多个特征。一个或多个全局变量基于目标内容的多个帧共有的一个或多个特征。在实施方式中,潜在空间可以包括与一个或多个全局变量对应的全局密度模型和与一个或多个局部变量对应的局部密度模型。全局密度可以由下式定义:

[0102] $$p_\theta(f) = \prod_{i}^{\dim(f)} p_\theta(f^i) * \mathcal{U}\left(-\frac{1}{2}, \frac{1}{2}\right)$$

[0103] 其中p_θ可以表示密度模型,f可以表示一个或多个全局变量,i可以表示与一个或多个全局变量的维度对应的维度索引,并且$\mathcal{U}(*)$可以表示均匀概率分布。应理解,可以使用其它值和/或模型来确定全局密度。静态密度$p_\theta(f^i)$可以由灵活的非参数、完全分解的模型参数化。累积概率密度可以由神经网络参数化,或者可以使用高斯概率分布。静态密度可以包括一个或多个非线性概率密度。

[0104] 局部密度模型可以由下式定义:

[0105] $$p_\theta(z_{1:T}) = \prod_{i}^{T} \prod_{i}^{\dim(z)} p_\theta(z_t^i | c_t) * \mathcal{U}\left(-\frac{1}{2}, \frac{1}{2}\right)$$

[0106] 其中p_θ可以表示密度模型,z可以表示一个或多个局部变量,T可以表示对应于给定帧的时间,i可以表示与一个或多个全局变量的维度对应的维度索引,c_t可以表示对应于T的上下文参数,并且$\mathcal{U}(*)$可以表示均匀概率分布。应理解,可以使用其它值和/或模型来确定局部密度。在一些实施方式中,$p_\theta(z_t^i|c_t) \equiv p_\theta(z_t^i|z_{<t})$。在实施方式中,$p_\theta(z_t^i|c_t) \equiv p_\theta(z_t^i|z_{t-1})$。

[0107] 在操作216处,可以生成多个分布。多个分布可以指示一个或多个局部变量和/或一个或多个全局变量的值的似然。对应于潜在空间的多个分布可以以一个或多个全局变量和一个或多个局部变量的平均值为中心。在一些实施方式中,可以在训练期间添加随机噪声,以近似在应用期间(例如,在训练之后)可能发生的舍入误差。例如,随机噪声可以是$\epsilon_i \sim \mathcal{U}\left(-\frac{1}{2}, \frac{1}{2}\right)$。应理解,随机噪声的量可以基于分布的生成方式。例如,分布可以由下式定义:

[0108] $$q_\phi(z_{1:T}, f \mid x_{1:T}) = q_\phi(f \mid x_{1:T}) \prod_{t=1}^{T} q_\phi(z_t \mid x_t)$$

[0109] 其中$q_\phi(z_{1:T}, f \mid x_{1:T})$可以表示真实后验的近似,φ可以表示一个或多个参数,z

可以表示一个或多个局部变量,T可以表示对应于给定帧的时间,并且f可以表示一个或多个局部变量。应理解,可以在不同的应用中使用不同的等式和变量。

[0110] 在一个示例中,一个或多个全局变量的均匀分布可以由下式定义:

[0111] $$\tilde{f}\sim q_{\phi}(f|x_{1:T})=\mathcal{U}\left(\hat{f}-\frac{1}{2},\hat{f}+\frac{1}{2}\right)$$

[0112] 其中$\tilde{f}$可以表示具有添加的噪声的潜在变量,q_{φ}可以表示全局后验的近似,f可以表示一个或多个全局变量,$x_{1:T}$可以表示时间T的目标内容的给定帧,$\mathcal{U}$可以表示添加的噪声,并且$\hat{f}$可以表示一个或多个全局变量的平均值。在一些实施方式中,$\hat{f}=\mu_{\phi}(x_{1:T})$。

[0113] 一个或多个局部变量的均匀分布可以由下式定义:

[0114] $$\tilde{z_t}\sim q_{\phi}(z_t|x_t)=\mathcal{U}\left(\hat{z_t}-\frac{1}{2},\hat{z_t}+\frac{1}{2}\right)$$

[0115] 其中$\tilde{z_t}$可以表示有噪声的局部潜在变量(例如,在添加噪声后的z),q_{φ}可以表示全局后验的近似,f可以表示一个或多个全局变量,$x_{1:T}$可以表示时间T的目标内容的给定帧,$\mathcal{U}$可以表示添加的噪声,并且$\hat{f}$可以表示一个或多个全局变量的平均值。在一些实施方式中,$\hat{z_t}=\mu_{\phi}(x_t)$。

[0116] 在操作218处,可以量化潜在空间。量化可以基于多个分布。在实施方式中,可以基于超过对应阈值来选择多个分布。例如,阈值可以基于以上等式中的一个或多个。量化可以将剩余值舍入到不同程度(例如,1000、100、10、1、0.1、0.01、0.001等)。

[0117] 在操作220处,可以对量化的潜在空间进行编码。编码可以包括熵编码以合并多个分布。熵编码可能导致二进制格式。通过利用来自预测模型的概率分布,可以将量化的潜在变量映射到简短的二进制描述中,如本文所述。熵编码的方法可以包括算术编码、范围编码、霍夫曼编码等。通过用二进制解码器对二进制进行解码,可以将二进制描述转换回到潜在变量,二进制解码器也可以包括概率分布。在一些实施方式中,概率分布可能已经由二进制解码器学习。应理解,其它无损形式的编码可以用于不同的应用。

[0118] 在操作222处,可以对编码的潜在空间进行解码。解码可以对称地跟随编码过程。在实施方式中,解码可以包括对编码的潜在空间进行熵解码。解码可以包括将熵解码的潜在空间与多层感知器模型组合。在一些实施方式中,解码可以包括将多个反卷积应用于熵解码的潜在空间与多层感知器模型、反卷积神经网络模型、转置卷积网络和/或其它模型的组合。

[0119] 上述公开技术的一个示例可以使用以下模型、函数和/或算法。$x_{1:T}=(x_1,\cdots,x_T)$可以表示视频序列T连续帧,其中T可以是大约10到大约100帧,但应理解,时间间隔可以更长或更短。生成模型可以捕获视频帧的概率分布,并允许生成可能的视频序列。在当前公开的技术中,可以使用两组潜在变量$z_{1:T}$和f。视频可以非线性地转换成潜在表示并存储为$z_{1:T}$和

f。生成模型可以包括视频和潜在变量上的概率分布：

[0120] $$p_\theta(x_{1:T}, z_{1:T}, f) = p_\theta(f)\prod_{t=1}^{T} p_\theta(z_t \mid z_{<t}) p_\theta(x_t \mid z_t, f)$$

[0121] 其中时间t处的帧x_t可以取决于对应的潜在变量z_t和f，并且θ是生成模型的所有参数的简写符号。先验分布$p_\theta(\tilde{f})$和$p_\theta(z_t|z_{<t})$可以通过各种技术来近似。在一些实施方式中，帧似然$p_\theta(x_{1:T}, z_{1:T}, f)$可以是拉普拉斯分布$\nabla(\mu_\theta(z_t, f), \lambda^{-1}1)$。对于$p_\theta(\tilde{f})$，可以使用概率密度的一般参数化。对于$p_\theta(z_t|z_{<t})$，可以使用上述几种模型，例如卡尔曼滤波器、LSTM和双向LSTM。条件网络可以通过使用与上述变量相同的定义，获取最有可能的帧$\tilde{x}_t = argmax\, p_\theta(x_t|f, z_t) = \mu_\theta(z_t, f)$，从而生成重构帧或解码的帧。

[0122] 可以从完全生成模型的条件部分获得解码器，该解码器从潜在变量中恢复当前帧的近似。解码的帧可以表示为$\hat{x}_t = p_\theta(x_t \mid f, z_t)$，其中$p_\theta(x_t|f, z_t)$可以被称为解码器。来自当前帧$x_t$的潜在变量可以由编码器q推断：

[0123] $$q_\phi(z_{1:T}, f \mid x_{1:T}) = q_\phi(f \mid x_{1:T}) \prod_{t=1}^{T} q_\phi(z_t \mid x_t)$$。

[0124] $q_\phi(z_{1:T}, f \mid x_{1:T})$可以是对真实后验$p_\phi(z_{1:T}, f \mid x_{1:T})$的变分近似。注意，f可以从序列中的所有视频帧中推断，而z_t可以从单个帧x_t中推断。换句话说，f可以包括对于段中的多个视频帧可以是共有的全局信息，并且zt可以包括对于每个帧实质上不同的本地信息或信息内容。编码器、解码器和f和z上的先验概率分布可以通过优化损失函数来共同学习。应理解，可以在不同的应用中使用不同的等式和变量。

[0125] 在一个示例中，对于64乘64像素的视频大小，解码器可以包括五个卷积层。对于层l＝1,2,3,4，步幅和填充的数量可以分别是2和1，并且卷积内核大小可以是4。步幅可以指滤波器如何围绕输入图像(input volume)卷积和/或滤波器可以如何在输入空间中移动。填充可以指添加空值来包围输入图像，以确保输出图像(output volume)与原始输入图像相匹配。层l＝1,2,3,4的信道数量可以是128、256、512、512。层5可以具有内核大小4、步幅1、填充0和信道数量1024。解码器架构可以被选择为与编码器对称，其中卷积层被反卷积(上采样)层代替。在实施方式中，f、z和h可以分别是512、64和1024。应理解，不同的值(例如，层数、步幅、填充、内核大小、信道数量等)可以用于不同的应用。

[0126] 应理解，尽管参考视频，但是当前公开的技术可以应用于其它媒体内容。

[0127] 图3A、图3B和图3C示出将当前公开的技术与现有技术进行比较的示例性压缩目标内容。如图所示，在时间t＝1时，图3A中的图像质量与H.265和VP9一样好或比其更好。在时间t＝6时，当前公开的技术比H.265和VP9好得多。使用当前公开的技术，每像素比特(bpp)约为0.06，PSNR约为44.6dB。H.265具有约0.86的bpp和约21.1dB的PSNR，并且VP9具有约0.57的bpp和约26.0dB的PSNR。

[0128] 图4A和图4B示出根据本公开的各种实施方式的将当前公开的技术与现有技术进

行比较的示例性压缩目标内容。如图所示,与现有技术诸如VP9相比,图像质量得到了改善。在图4A中,当前公开的技术具有约0.29的bpp和约38.1的PSNR,相比之下VP9的bpp约为0.44、PSNR约为25.7。在当前公开的技术中,图像也比图4A中的VP9更清晰。在图4B中,当前公开的技术针对顶部图像具有约0.39的bpp和约32.0的PSNR并且针对底部图像具有约30.1的PSNR。VP9针对顶部图像具有约0.39的bpp和约29.3的PSNR并且针对底部图像具有约30.8的PSNR。

[0129] 图5A、图5B、图6A和图6B示出所公开的技术在视频数据集方面的示例性能。图5A和图5B可以使用第一数据集,图6A和图6B可以使用第二数据集。每个视频集的大小为约64乘约64像素。如图所示,曲线502、504、506和508分别使用MPEG-4part2、H.264、H.265和VP9而示出。曲线510和512分别使用卡尔曼滤波器模型和LSTM模型而示出。x轴对应于比特/像素的比特率(较低表示性能更好),y轴对应于在PSNR或MS-SSIM中测量的失真(较高表示性能更好)。每个编解码器产生率失真曲线,更好的性能对应于高于其它编解码器的曲线。如图所示,在这些视频方面,当前公开的技术优于大多数编解码器。

[0130] 图7A、图7B和图7C示出根据本公开的各种实施方式的使用多个压缩模型的示例性率失真曲线。图7A可以使用第一训练内容,图7B可以使用第二训练内容,并且图7C可以使用第三训练内容。如图所示,较高的曲线表示较好的性能。曲线702、704和706分别示出H.264、H2.65和VP9,而曲线708、710和712示出当前公开技术的实施方式(例如,分别为KFP-LG、LSTMP-LG和LSTMP-L)。如上所述,轴基本上类似于图5A、图5B、图6A和图6B。

[0131] 图8A、图8B和图8C示出根据本公开的各种实施方式的使用多个压缩模型的示例性率失真曲线。图8A可以使用第一训练内容,图8B可以使用第二训练内容,并且图8C可以使用第三训练内容。x轴可以表示每个像素的比特或比特率,y轴可以表示多尺度结构相似度值,其可以是近似于结构形成的感知变化的基于感知的度量。较高的曲线可以表示较小的失真。曲线802、804和806分别示出H.264、H2.65和VP9,而曲线808、810和812示出当前公开技术的实施方式(例如,分别为KFP-LG、LSTMP-LG和LSTMP-L)。

[0132] 图9A、图9B和图9C示出根据本公开的各种实施方式的使用多个压缩模型的示例性信息平均比特。图9A可以使用第一训练内容,图9B可以使用第二训练内容,并且图9C可以使用第三训练内容。x轴可以表示帧索引,或者给定的帧或段,y轴可以表示存储在f和z中的信息平均比特。在图9A中,PSNR约为43.2;在图9B中,PSNR约为37.1;在图9C中,PSNR约为30.3。当多个压缩模型适应视频序列时,熵随帧索引下降。曲线902、904和906示出当前公开技术的实施方式(例如,分别为LSTMP-L、KFP-LG和LSTMP-LG)。

[0133] 图10A、图10B、图10C和图10D示出根据本公开的各种实施方式的使用当前公开的技术的示例性分布。如图所示,可以将先验分布与压缩模型的后验的经验分布进行比较。图10A和图10B示出给定全局变量的二维。图10C和10D示出给定局部变量的二维。x轴表示观察结果,y轴表示观察结果的概率。更高的条形图表示观察结果的概率更大。如图所示,压缩模型的后验是围绕表示先验分布的内部条形1004的较大条形1002。条形1002基本上与图10A、图10B、图10C和图10D中的条形1004匹配。

[0134] 如本文所使用的,术语部件可以描述可以根据本文公开技术的一个或多个实施方式执行的给定功能单元。如本文所使用的,部件可以利用任何形式的硬件、软件或其组合来实现。例如,可以实现一个或多个处理器、控制器、ASIC、PLA、PAL、CPLD、FPGA、逻辑部件、软

件例程或其它机制来组成部件。在实施方式中,本文描述的各种部件可以实现为分立部件,或者所描述的功能和特征可以在一个或多个部件之间部分或全部共享。换句话说,在阅读本说明书之后,对于本领域普通技术人员明显的是,本文描述的各种特征和功能可以在任何给定的应用中实现,并且可以以各种组合和排列方式在一个或多个单独或共享的部件中实现。如本文所使用的,术语引擎可以描述经配置为执行一个或多个特定任务的部件的集合。即使功能的各种特征或元件可以被单独描述或声明为单独的部件或引擎,但是本领域普通技术人员将理解,这些特征和功能可以在一个或多个通用软件和硬件元件之间共享,并且此描述不要求或暗示使用单独的硬件或软件部件来实现此类特征或功能。

[0135] 在使用软件全部或部分地实现引擎、部件或本技术的部件的情况下,在一个实施方式中,这些软件元件可以被实现为与能够执行关于其描述的功能的计算或处理部件一起操作。在图11中示出一个这样的示例性计算部件。根据该示例性计算部件1100描述各种实施方式。在阅读本说明书之后,相关领域的技术人员将明白如何使用其它计算部件或架构来实现该技术。

[0136] 现在参考图11,计算部件1100可以表示例如在台式机、膝上型计算机和笔记本计算机;手持计算设备(PDA、智能手机、手机、掌上电脑等);大型机、超级计算机、工作站或服务器;或者对于给定的应用或环境可能需要或合适的任何其它类型的专用或通用计算设备中发现的计算或处理能力。计算部件1100还可以表示嵌入给定设备中或对于给定设备以其他方式可用的计算能力。例如,计算部件可以在其它电子设备中找到,其它电子设备例如为数码相机、导航系统、蜂窝电话、便携式计算设备、调制解调器、路由器、WAP、终端和可能包括某种形式的处理能力的其它电子设备。

[0137] 计算部件1100可以包括例如一个或多个处理器、物理计算机处理器控制器、控制部件或其它处理设备,诸如处理器1104。可以使用通用或专用处理引擎例如微处理器、控制器或其它控制逻辑来实现处理器1104。在所示的示例中,处理器1104连接到总线1102,但是任何通信介质都可以用于促进与计算部件1100的其它部件的交互或外部通信。

[0138] 计算部件1100还可以包括一个或多个存储器部件,本文简称为主存储器1108。例如,优选地,随机存取存储器(RAM)或其它动态存储器可以用于存储将由处理器1104执行的信息和指令。主存储器1108还可以用于在执行由处理器1104执行的指令期间存储临时变量或其它中间信息。计算部件1100同样可以包括耦合到总线1102的只读存储器(“ROM”)或其它静态存储设备,以用于为处理器1104存储静态信息和指令。

[0139] 计算部件1100还可以包括一种或多种不同形式的信息存储设备1110,其可以包括例如媒体驱动器1112和存储单元接口1120。媒体驱动器1112可以包括驱动器或其它机构,以支持固定或可移动的存储介质1114。例如,可以提供硬盘驱动器、软盘驱动器、磁带驱动器、光盘驱动器、CD或DVD驱动器(R或RW)、非瞬态电子存储设备和/或其它可移动或固定的媒体驱动器。因此,存储介质1114可以包括例如硬盘、软盘、磁带、盒式磁带、光盘、CD或DVD或由媒体驱动器1112读取、写入或访问的其它固定或可移动的介质。如这些示例所示,存储介质1114可以包括其中存储有计算机软件或数据的计算机可用存储介质。

[0140] 在替代实施方式中,信息存储机构1110可以包括用于允许将计算机程序或其它指令或数据加载到计算部件1100中的其它类似媒介。此类媒介可以包括例如固定或可移动的存储单元1122和接口1120。此类存储单元1122和接口1120的示例可以包括程序盒和盒接

口、可移动存储器(例如,闪存或其它可移动存储器部件)和存储器插槽、PCMCIA插槽和卡、以及允许将软件和数据从存储单元1122传输到计算部件1100的其它固定或可移动的存储单元1122和接口1120。

[0141] 计算部件1100也可以包括通信接口1124。通信接口1124可以用于允许软件和数据在计算部件1100与外部设备之间传输。通信接口1124的示例可以包括调制解调器或软调制解调器、网络接口(诸如以太网、网络接口卡、WiMedia、IEEE802.XX或其它接口)、通信端口(诸如USB端口、IR端口、RS232端口、Bluetooth®接口或其它端口)或其它通信接口。经由通信接口1124传输的软件和数据可以承载在信号上,信号可以是电子的、电磁的(包括光学的),或者能够由给定通信接口1124交换的其它信号。这些信号可以经由信道1128提供给通信接口1124。该信道1128可以承载信号,并且可以使用有线或无线通信介质来实现。信道的一些示例可以包括电话线、蜂窝链路、RF链路、光链路、网络接口、局域网或广域网以及其它有线或无线通信信道。

[0142] 在本文件中,术语“计算机程序介质”和“计算机可用介质”通常用于指代介质,例如存储器1108、存储单元1120、介质1114和信道1128。这些和其它各种形式的计算机程序介质或计算机可用介质可以涉及将一个或多个指令的一个或多个序列携带到处理设备以供执行。在介质上体现的此类指令通常被称为“计算机程序代码”或“计算机程序产品”(其可以以计算机程序或其它分组的形式分组)。当被执行时,此类指令可以使计算部件1100能够执行本文讨论的公开技术的特征或功能。

[0143] 虽然上面已经描述了公开技术的各种实施方式,但是应理解,它们仅以示例而非限制的方式呈现。同样地,各种图可以描绘公开技术的示例性架构或其它配置,这是为了帮助理解可以包括在公开技术中的特征和功能。公开的技术不限于所示的示例性架构或配置,而是可以使用各种替代架构和配置来实现期望的特征。事实上,对于本领域技术人员来说,可以如何实现替代的功能、逻辑或物理分区和配置以实现本文公开的技术的期望特征是明显的。此外,除了本文描述的那些之外的多个不同的组成部件名称可以应用于各种分区。另外,关于流程图、操作描述和方法权利要求,除非上下文另有说明,否则本文呈现的步骤的顺序不应强制实施各种实施方式以按相同的顺序执行所述功能。

[0144] 尽管以上根据各种示例性实施方式和实施方式描述了公开的技术,但是应理解,在一个或多个单独实施方式中描述的各种特征、方面和功能不限于它们对与其一起描述的特定实施方式的适用性,而是相反可以单独或以各种组合的方式应用于公开技术的其它实施方式中的一个或多个,无论是否描述了此类实施方式以及此类特征是否被呈现为所描述的实施方式的一部分。因此,本文公开的技术的广度和范围不应受任何上述示例性实施方式的限制。

[0145] 除非另有明确说明,否则本文件中使用的术语和短语及其变体应理解为开放式的,而不是限制性的。作为前述的示例:术语“包括(including)”应理解为“包括但不限于”等含义;术语“示例(example)”用于提供所讨论项目的示例性实例,而不是其穷举或限制性列表;术语“一(a)”或“一个(an)”应理解为“至少一个”、“一个或多个”等含义;并且形容词诸如“常规的(conventional)”、“传统的(traditional)”、“正常的(normal)”、“标准的(standard)”、“已知的(known)”以及类似含义的术语不应被解释为将所描述的项目限制到给定时间段或给定时间可用的项目,而是相反应被理解为包括现在或将来任何时候可用或

已知的常规、传统、正常或标准的技术。同样地,在本文件涉及对于本领域普通技术人员来说明显或已知的技术的情况下,此类技术包括那些现在或将来任何时候对于本领域技术人员来说明显或已知的技术。

[0146] 在某些情况下,诸如“一个或多个(one or more)”、“至少(at least)”、“但不限于(but not limited to)”或其它类似短语的扩大单词和短语的存在不应被理解为意味着在可能不存在此类扩大短语的情况下意图或需要更窄的情况。术语“部件(component)”的使用并不意味着作为部件的一部分描述或要求保护的部件或功能都配置在公共封装中。事实上,部件的任何或所有不同部件,无论是控制逻辑还是其它部件,都可以组合在单个封装中或单独维护,并且还可以分布在多个组或封装中或跨多个位置。

[0147] 另外,本文阐述的各种实施方式是根据示例性框图、流程图和其它图示来描述的。如在阅读本文件之后对于本领域的普通技术人员将变得明显的是,可以在不局限于所示示例的情况下实现所示实施方式及其各种替代方案。例如,框图及其附带的描述不应被解释为强制要求特定的架构或配置。

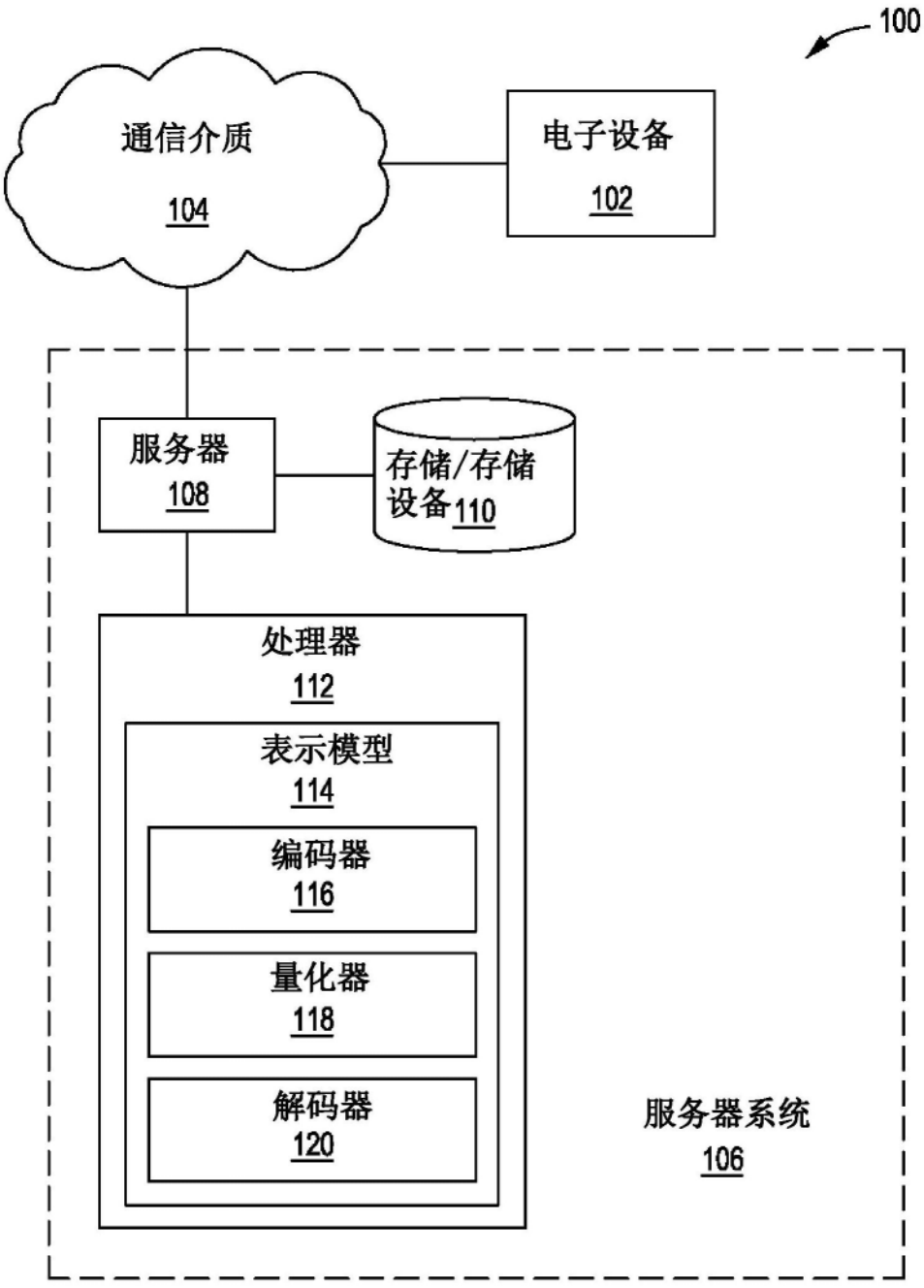

100
通信介质
104
电子设备
102
服务器
108
存储/存储设备110
处理器
112
表示模型
114
编码器
116
量化器
118
解码器
120
服务器系统
106

图1A

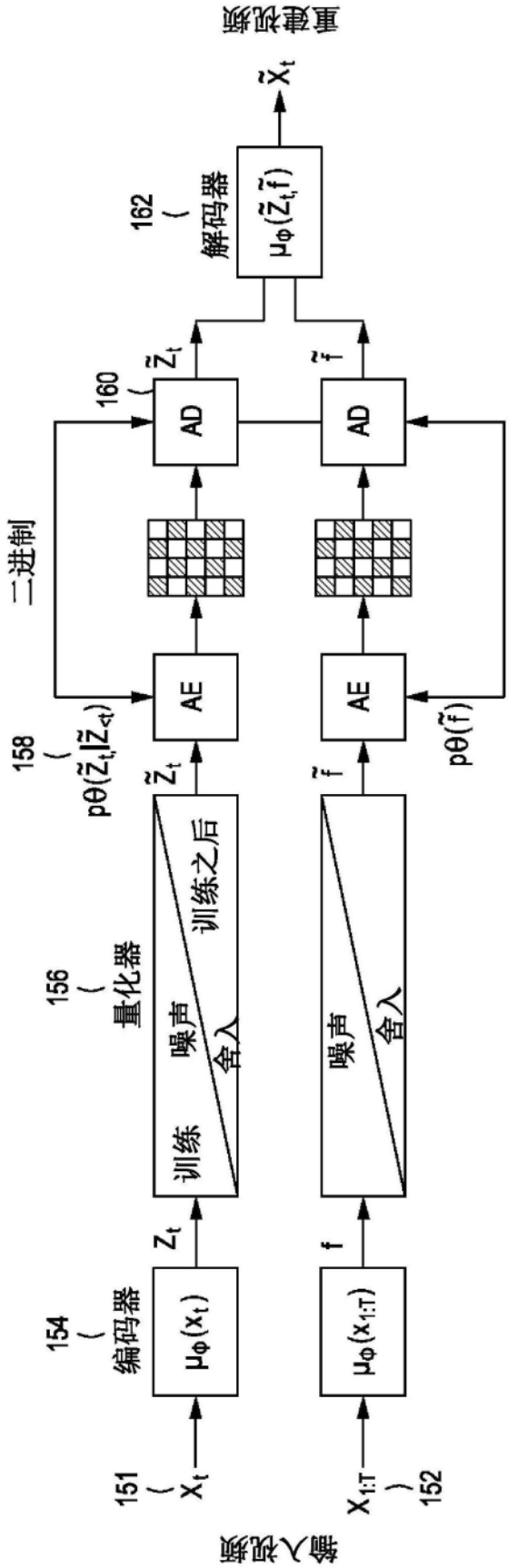

重建视频
$\tilde{x}_t$
162
解码器
$\mu_\phi(\tilde{z}_t,\tilde{f})$
160
AD
AD
$\tilde{z}_t$
$\tilde{f}$
二进制
158
$p\theta(\tilde{z}_t|\tilde{z}_{<t})$
AE
AE
$p\theta(\tilde{f})$
$\tilde{z}_t$
$\tilde{f}$
156
量化器
训练
噪声
舍入
训练之后
噪声
舍入
z_t
f
154
编码器
$\mu_\phi(x_t)$
$\mu_\phi(x_{1:T})$
151
x_t
$x_{1:T}$
152
输入视频

图1B

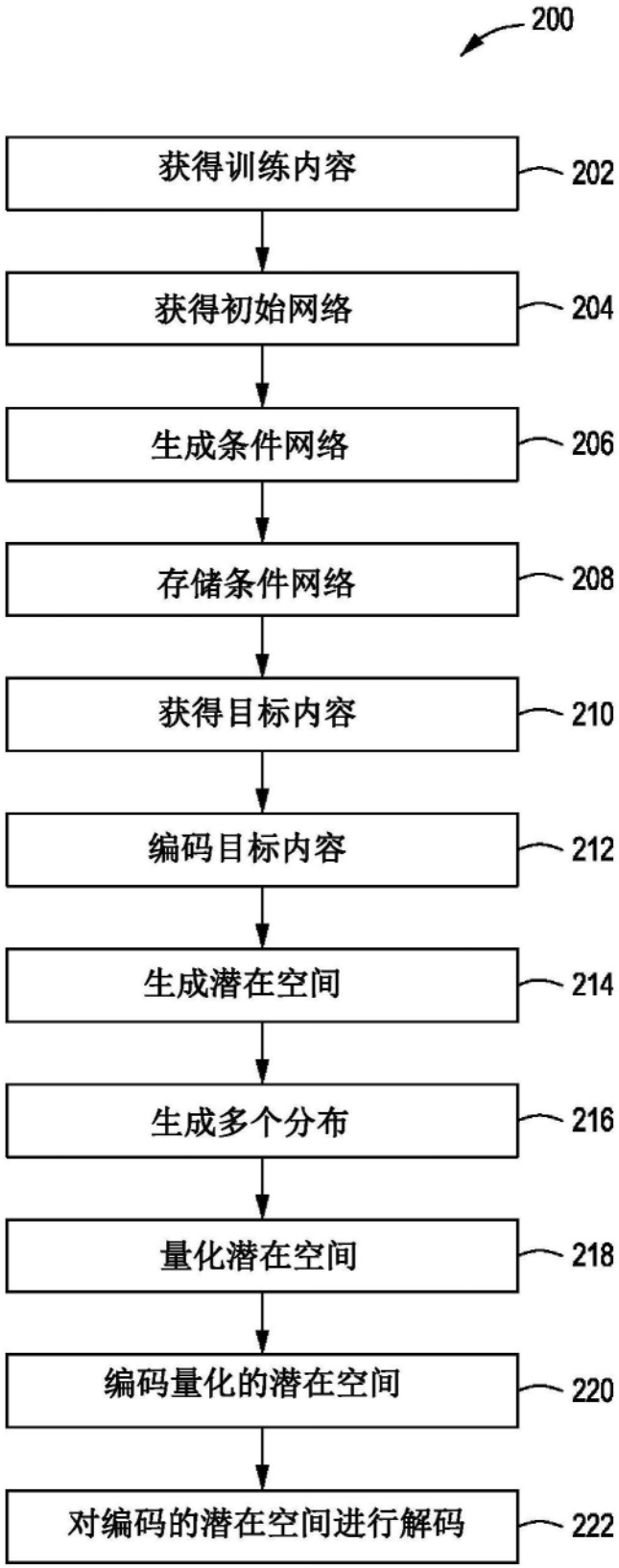

200
获得训练内容
202
获得初始网络
204
生成条件网络
206
存储条件网络
208
获得目标内容
210
编码目标内容
212
生成潜在空间
214
生成多个分布
216
量化潜在空间
218
编码量化的潜在空间
220
对编码的潜在空间进行解码
222

图2

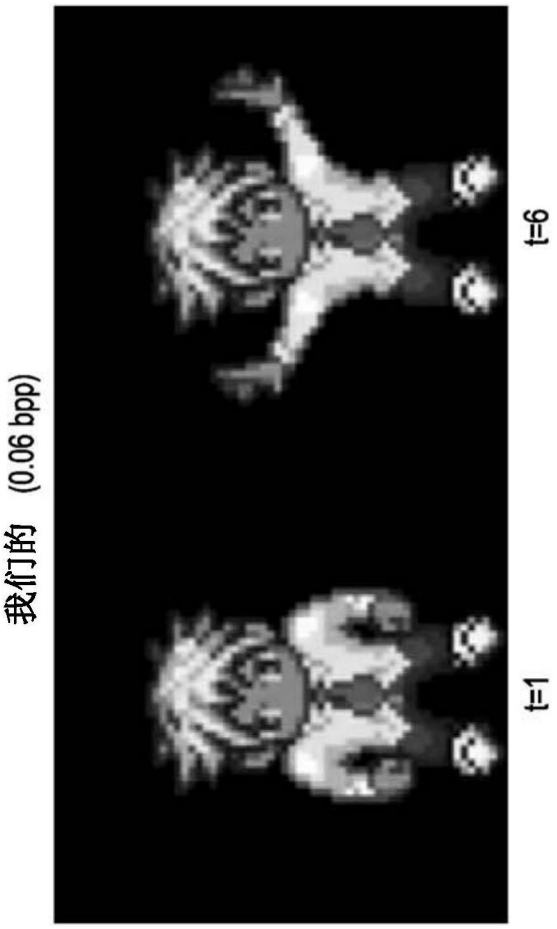


图3A

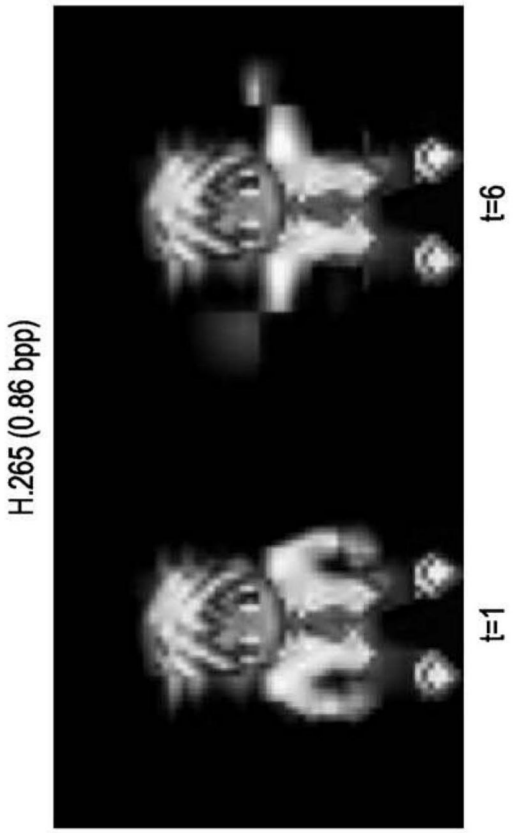


图3B

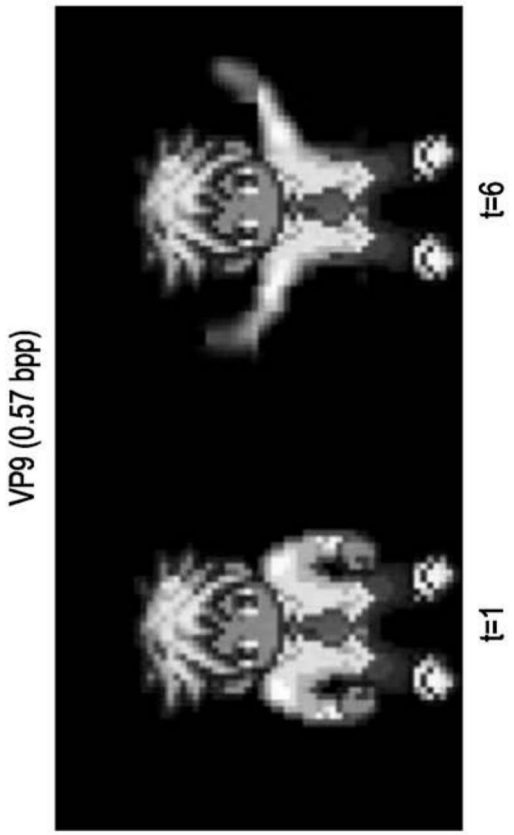


图3C

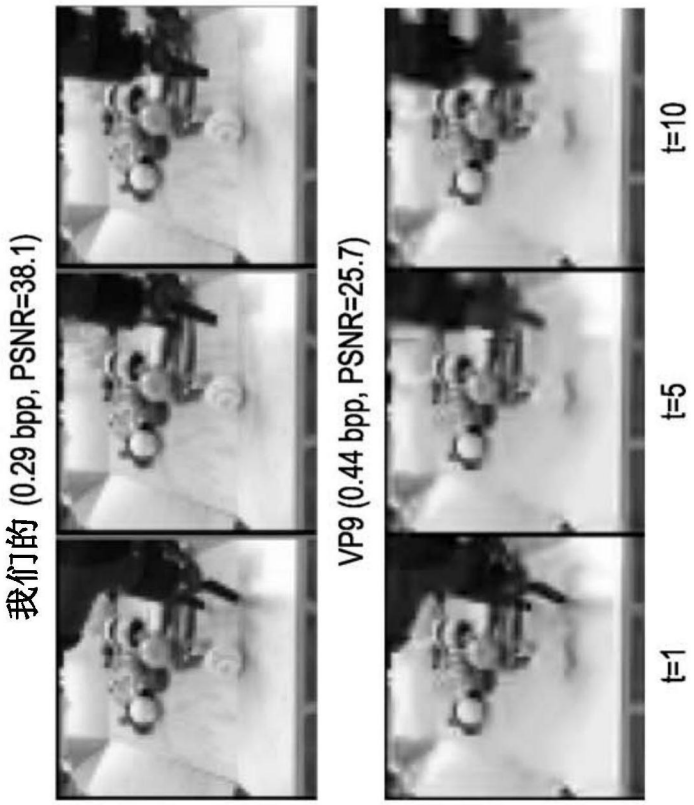


图4A

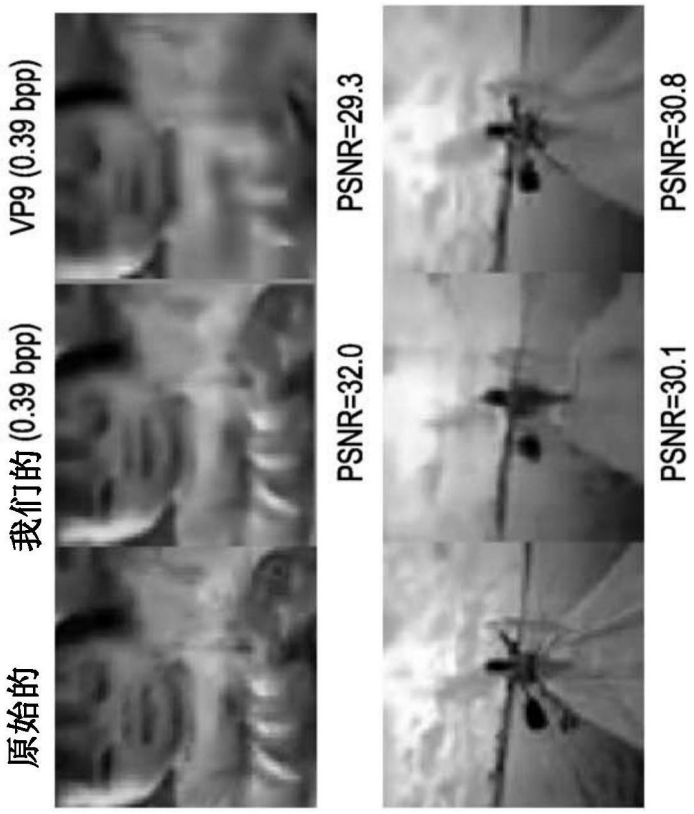


图4B

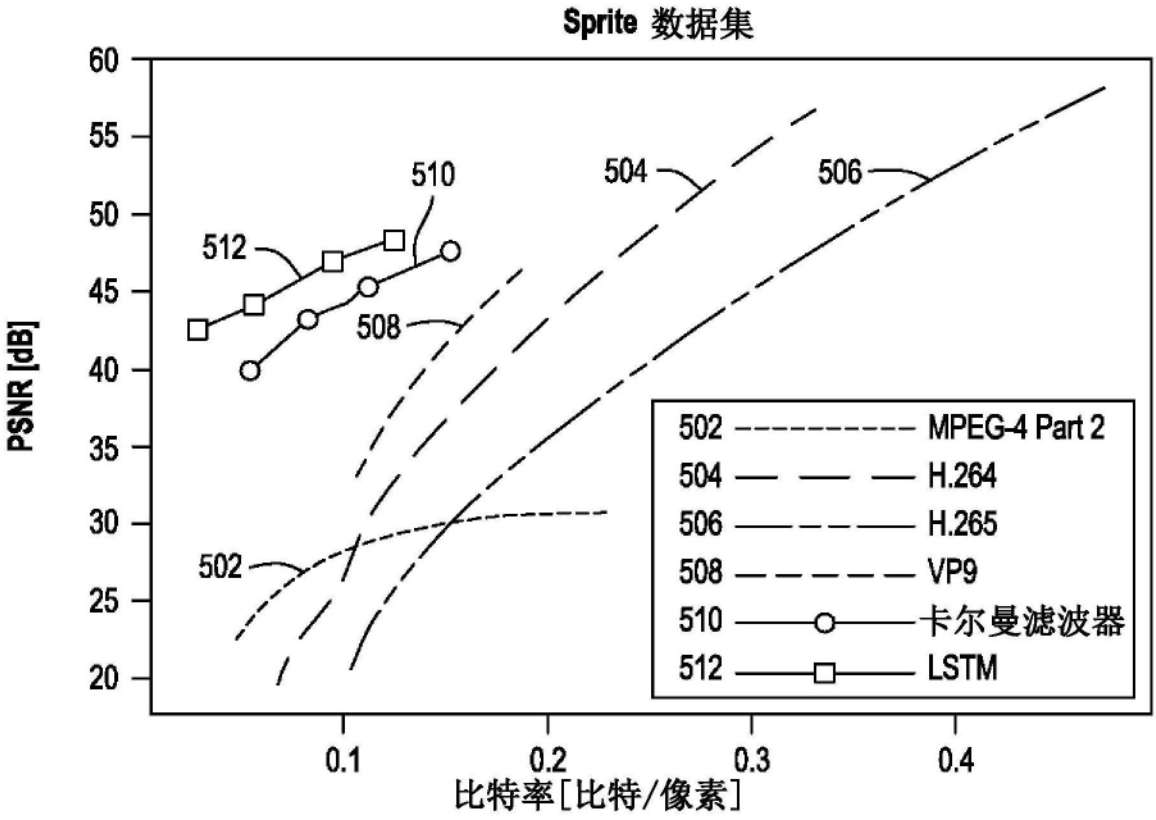

Sprite 数据集
PSNR [dB]
60
55
50
45
40
35
30
25
20
0.1
0.2
0.3
0.4
比特率[比特/像素]
502
504
506
508
510
512
MPEG-4 Part 2
H.264
H.265
VP9
卡尔曼滤波器
LSTM

图5A

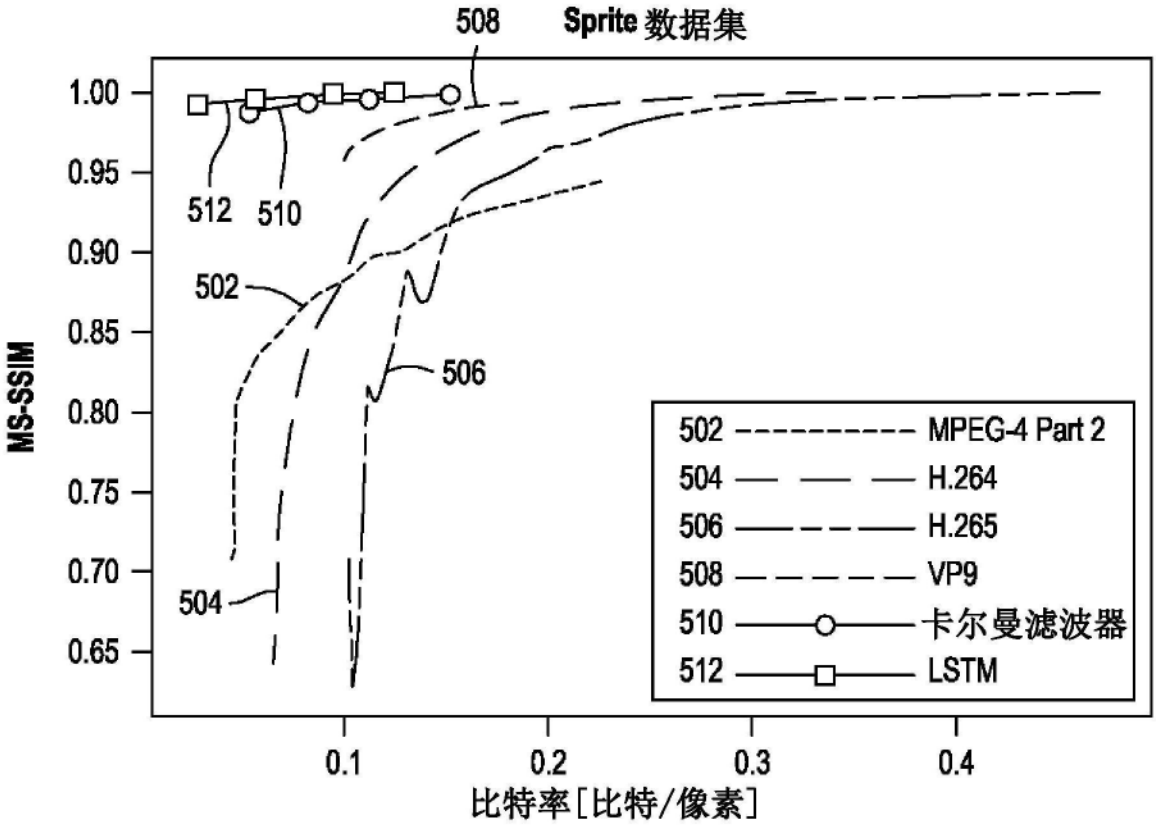

508
Sprite 数据集
1.00
0.95
0.90
0.85
0.80
0.75
0.70
0.65
MS-SSIM
512
510
502
506
504
0.1
0.2
0.3
0.4
比特率[比特/像素]
502 MPEG-4 Part 2
504 H.264
506 H.265
508 VP9
510 卡尔曼滤波器
512 LSTM

图5B

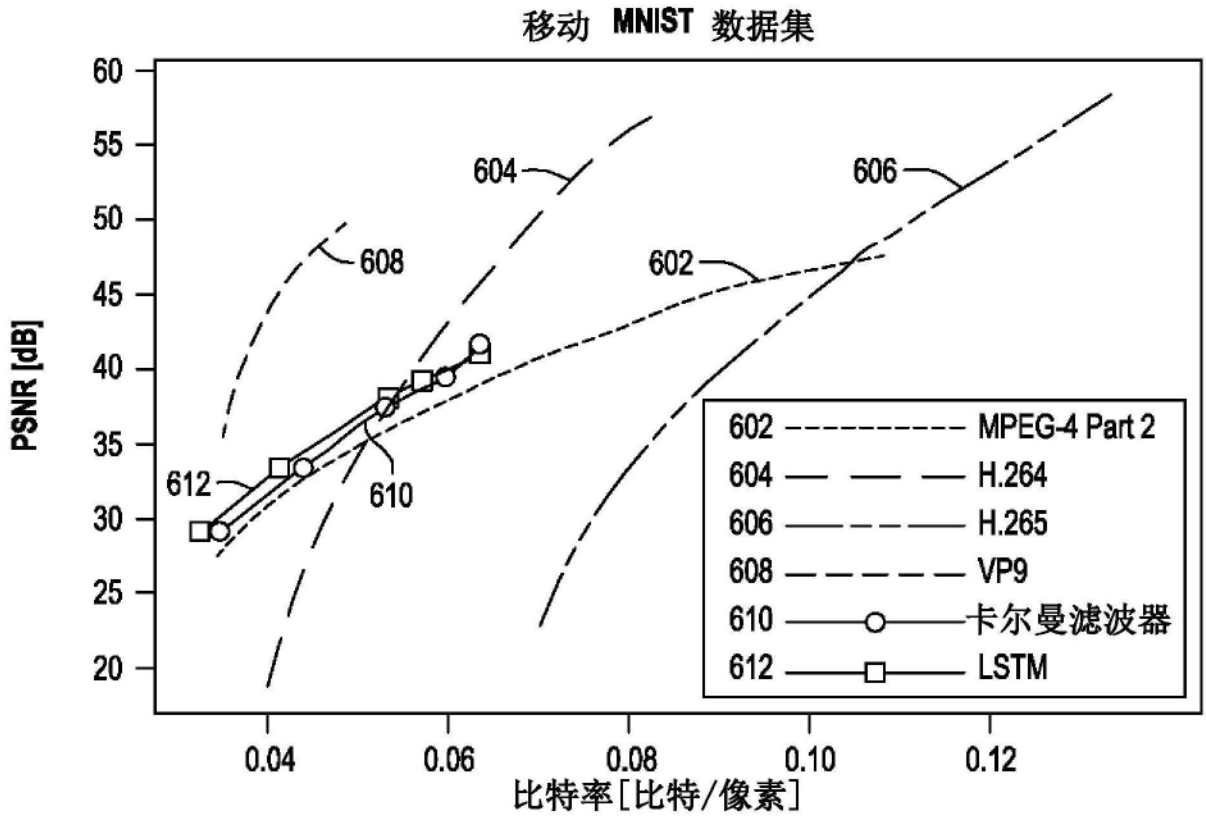

移动 MNIST 数据集
PSNR [dB]
60
55
50
45
40
35
30
25
20
0.04
0.06
0.08
0.10
0.12
比特率[比特/像素]
602
604
606
608
610
612
602 MPEG-4 Part 2
604 H.264
606 H.265
608 VP9
610 卡尔曼滤波器
612 LSTM

图6A

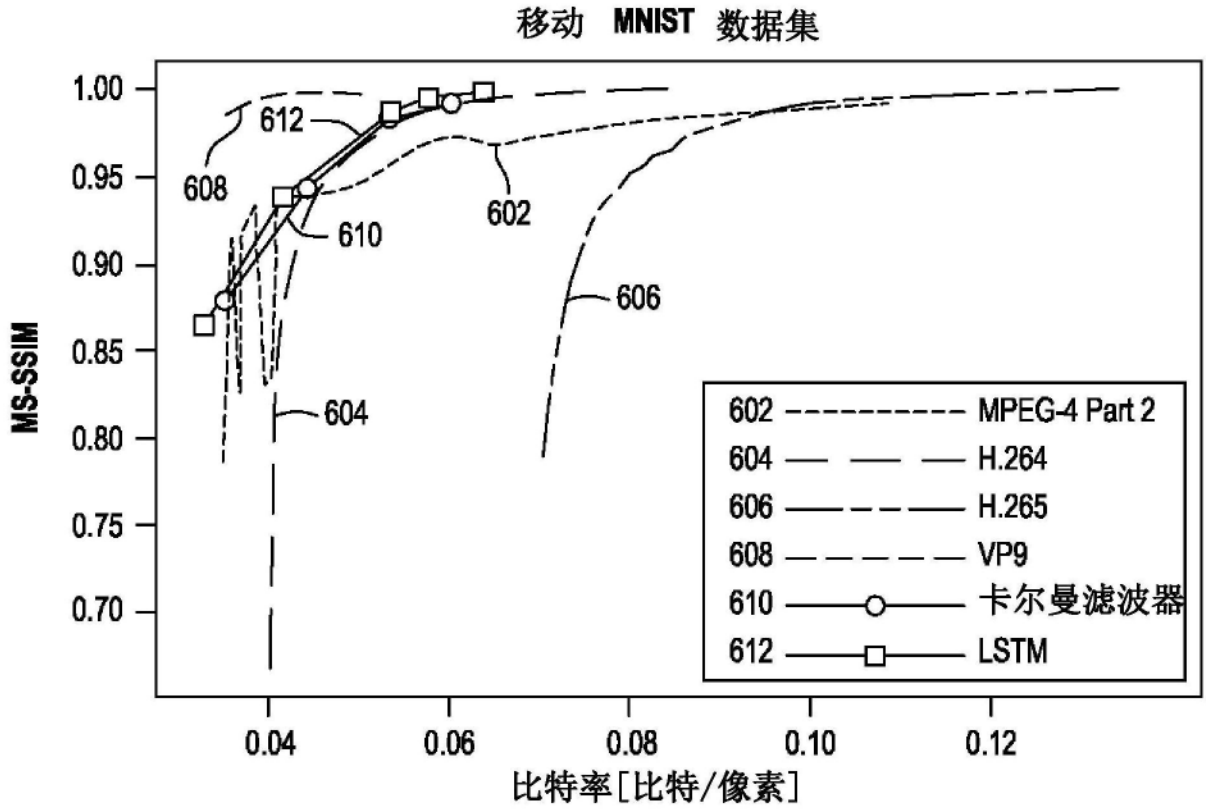

移动 MNIST 数据集
MS-SSIM
1.00
0.95
0.90
0.85
0.80
0.75
0.70
0.04
0.06
0.08
0.10
0.12
比特率[比特/像素]
602 MPEG-4 Part 2
604 H.264
606 H.265
608 VP9
610 卡尔曼滤波器
612 LSTM

图6B

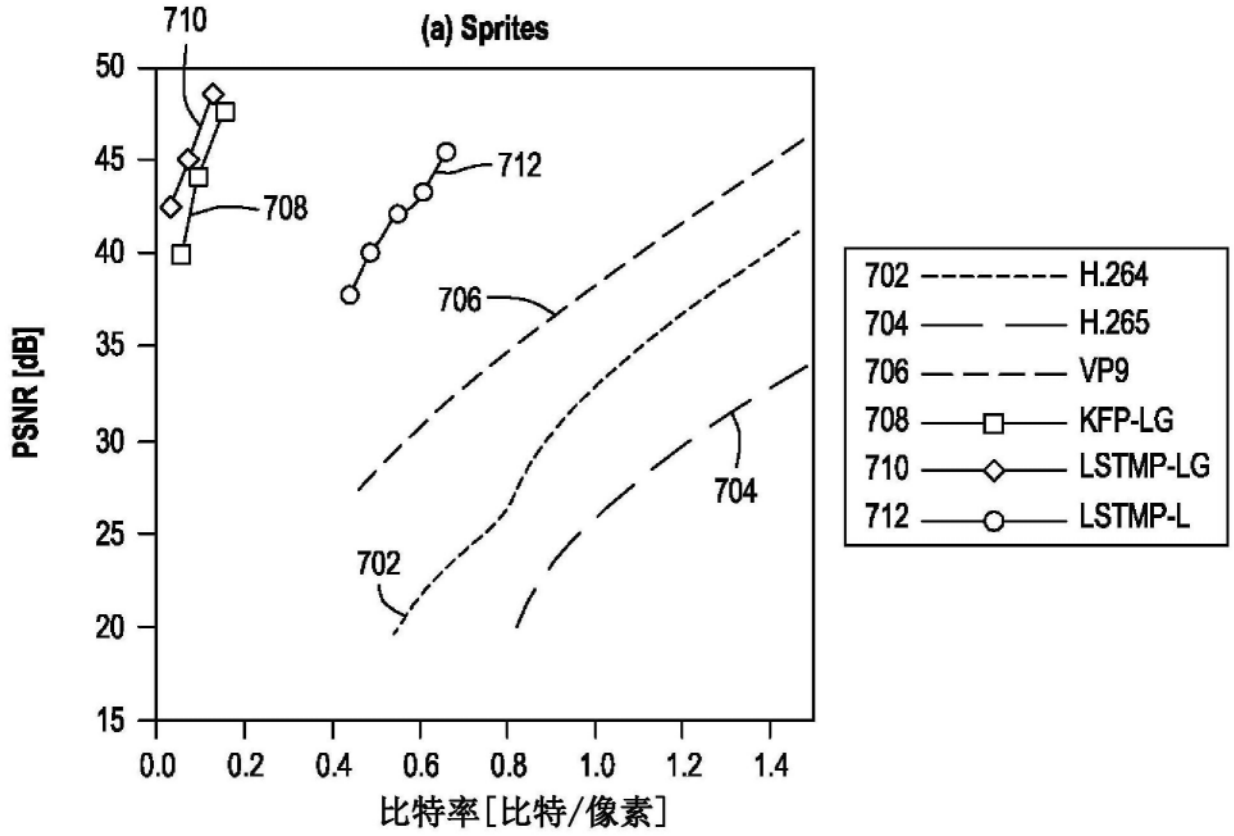

(a) Sprites
PSNR [dB]
比特率[比特/像素]
50
45
40
35
30
25
20
15
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
710
708
712
706
704
702
702 H.264
704 H.265
706 VP9
708 KFP-LG
710 LSTMP-LG
712 LSTMP-L

图7A

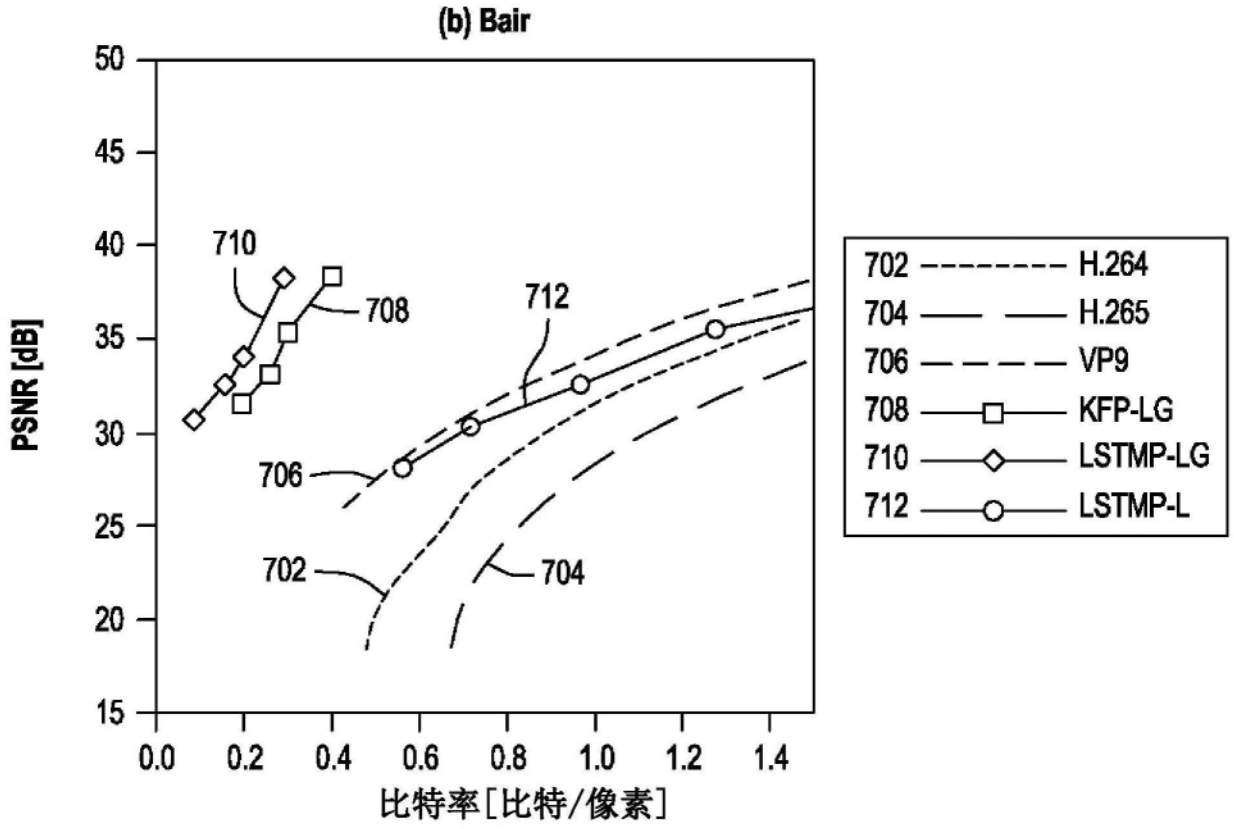

(b) Bair
PSNR [dB]
50
45
40
35
30
25
20
15
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
比特率[比特/像素]
710
708
712
706
702
704
702 H.264
704 H.265
706 VP9
708 KFP-LG
710 LSTMP-LG
712 LSTMP-L

图7B

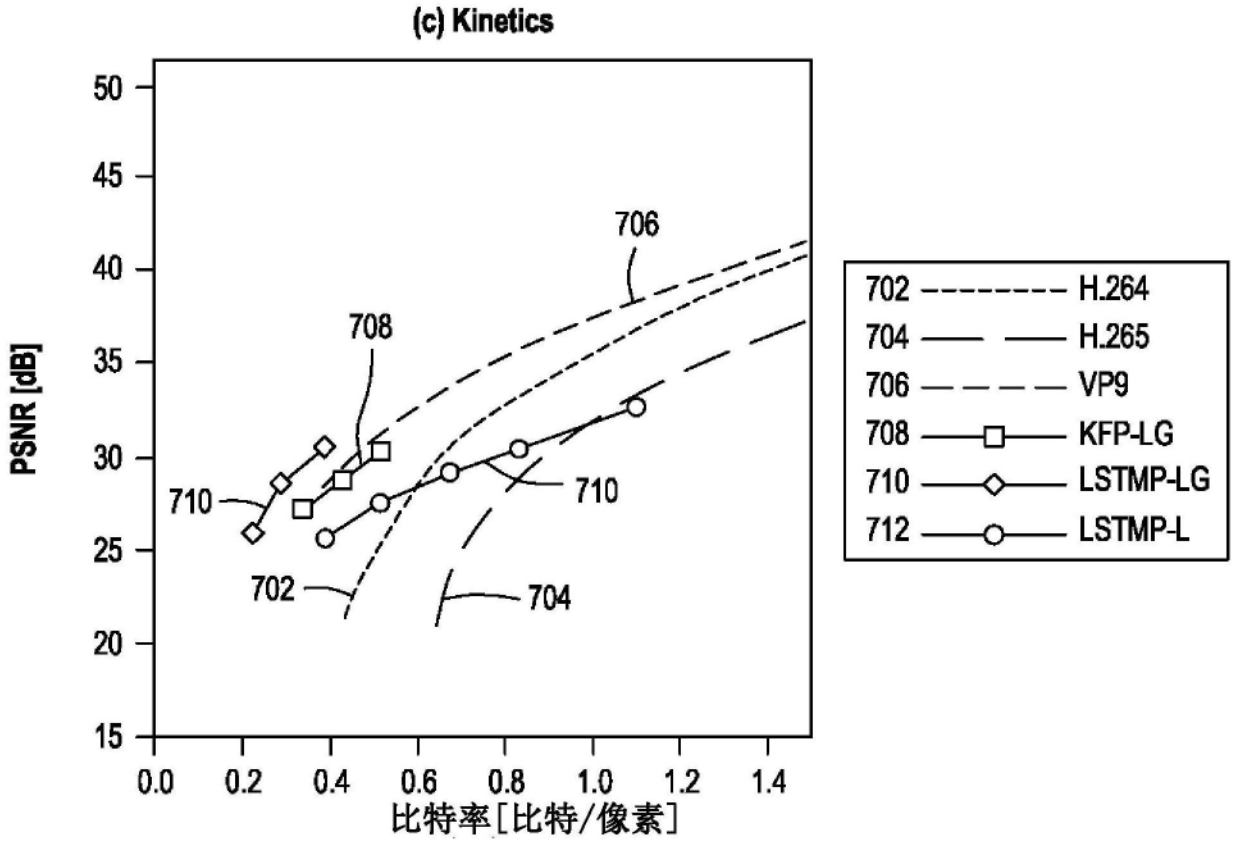

(c) Kinetics
PSNR [dB]
50
45
40
35
30
25
20
15
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
比特率[比特/像素]
706
708
710
710
702
704
702 H.264
704 H.265
706 VP9
708 KFP-LG
710 LSTMP-LG
712 LSTMP-L

图7C

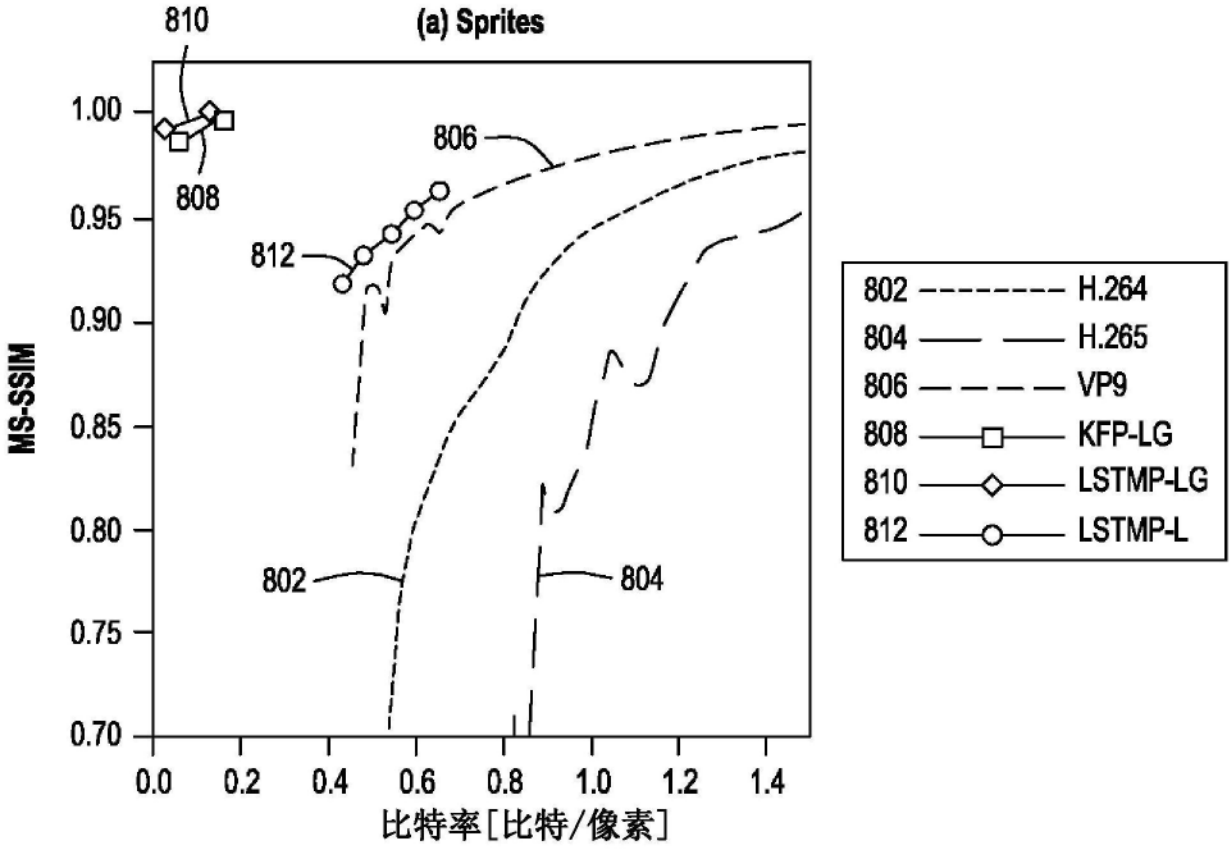

(a) Sprites
MS-SSIM
1.00
0.95
0.90
0.85
0.80
0.75
0.70
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
比特率[比特/像素]
810
808
806
812
802
804
802 H.264
804 H.265
806 VP9
808 KFP-LG
810 LSTMP-LG
812 LSTMP-L

图8A

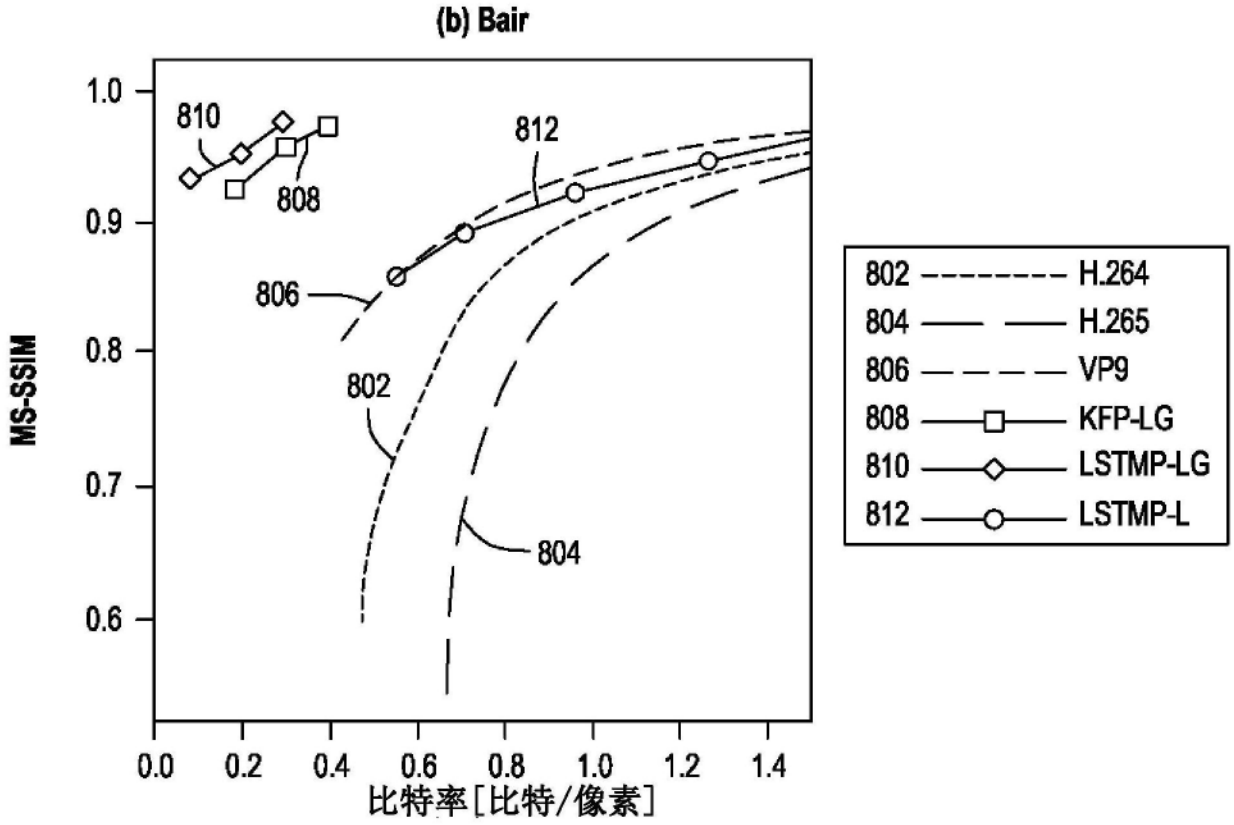

(b) Bair
MS-SSIM
1.0
0.9
0.8
0.7
0.6
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
比特率[比特/像素]
810
812
808
806
802
804
802 H.264
804 H.265
806 VP9
808 KFP-LG
810 LSTMP-LG
812 LSTMP-L

图8B

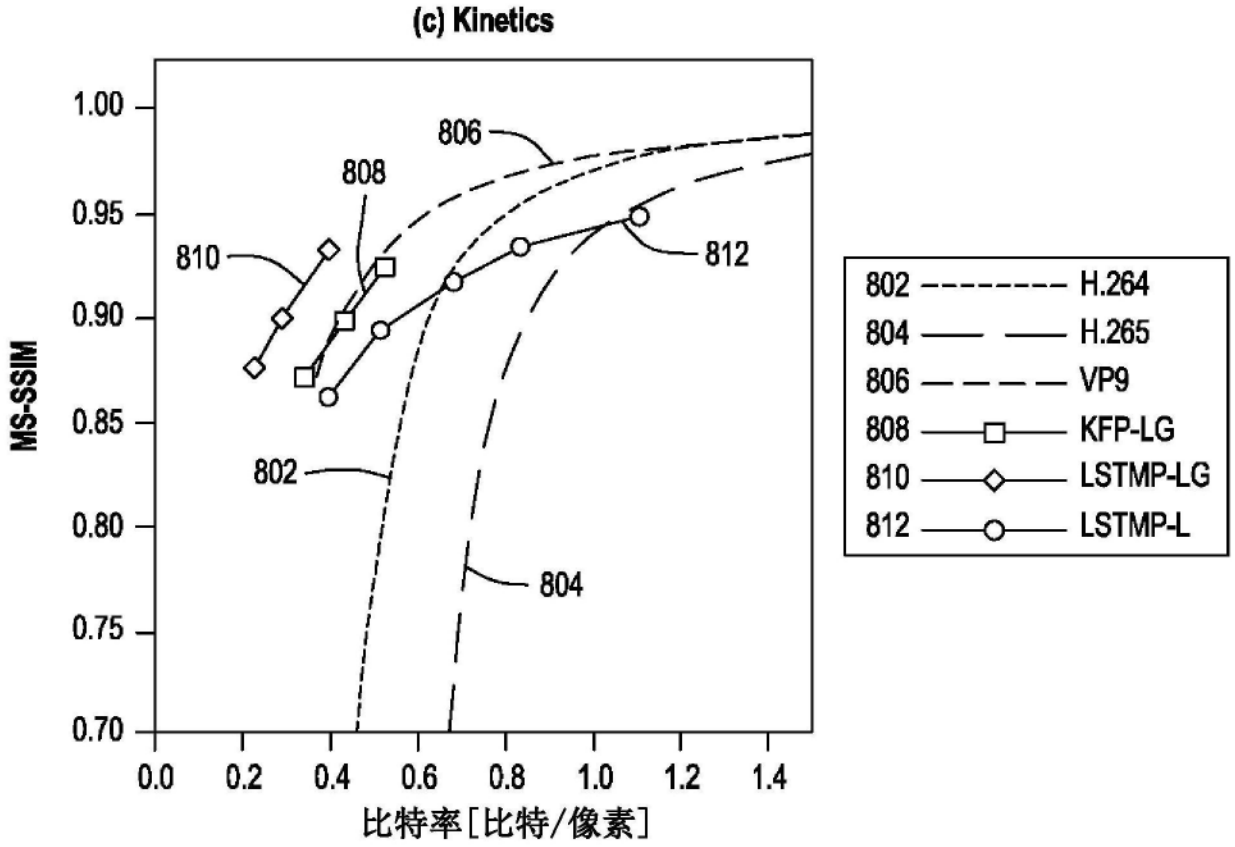

(c) Kinetics
MS-SSIM
1.00
0.95
0.90
0.85
0.80
0.75
0.70
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
比特率[比特/像素]
806
808
810
812
802
804
802 H.264
804 H.265
806 VP9
808 KFP-LG
810 LSTMP-LG
812 LSTMP-L

图8C

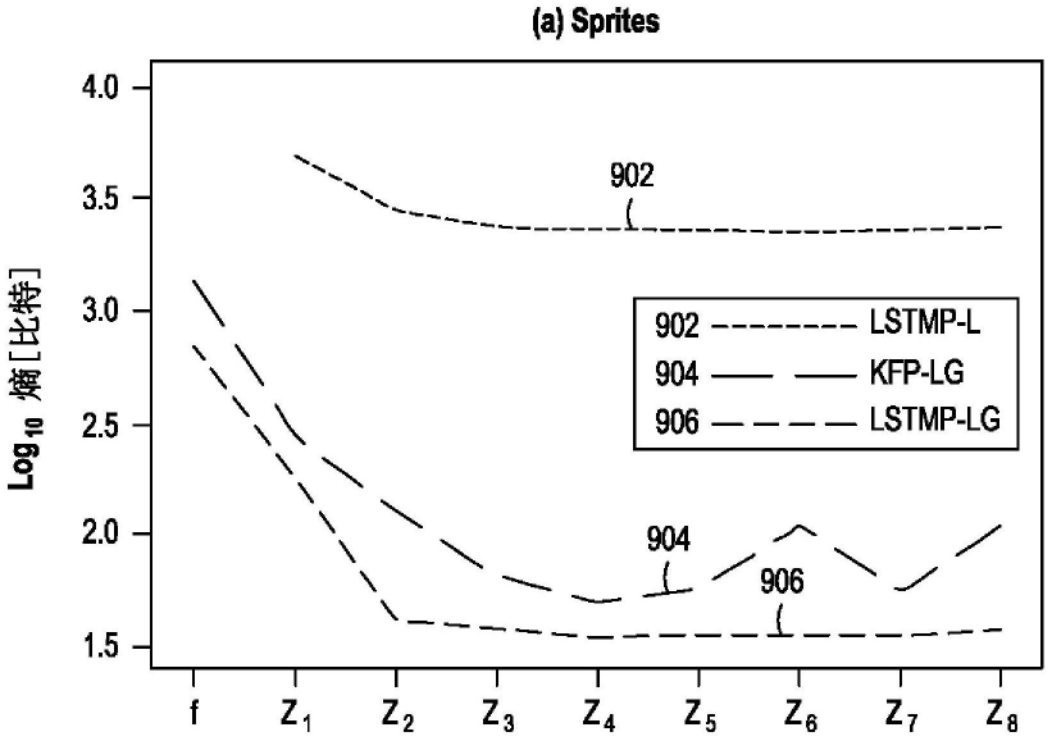

(a) Sprites
Log10 熵[比特]
4.0
3.5
3.0
2.5
2.0
1.5
f
Z1
Z2
Z3
Z4
Z5
Z6
Z7
Z8
902
904
906
902 LSTMP-L
904 KFP-LG
906 LSTMP-LG

图9A

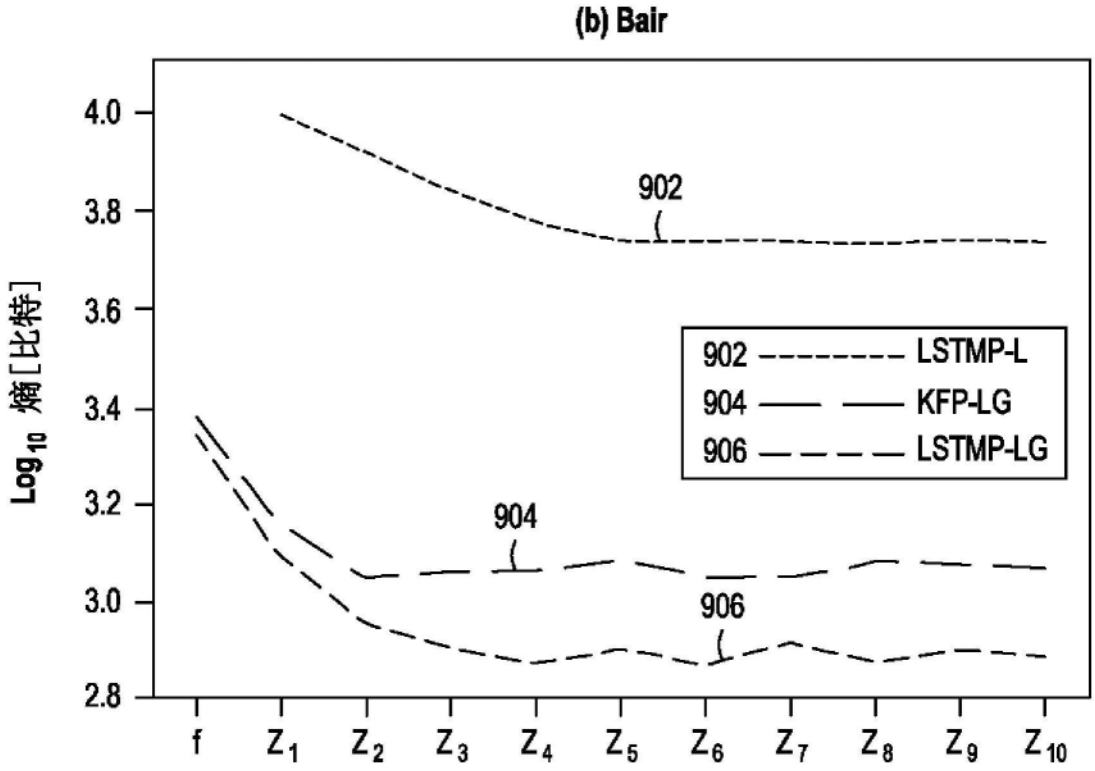

(b) Bair
4.0
3.8
3.6
3.4
3.2
3.0
2.8
Log10 熵[比特]
902
904
906
902 LSTMP-L
904 KFP-LG
906 LSTMP-LG
f
Z1
Z2
Z3
Z4
Z5
Z6
Z7
Z8
Z9
Z10

图9B

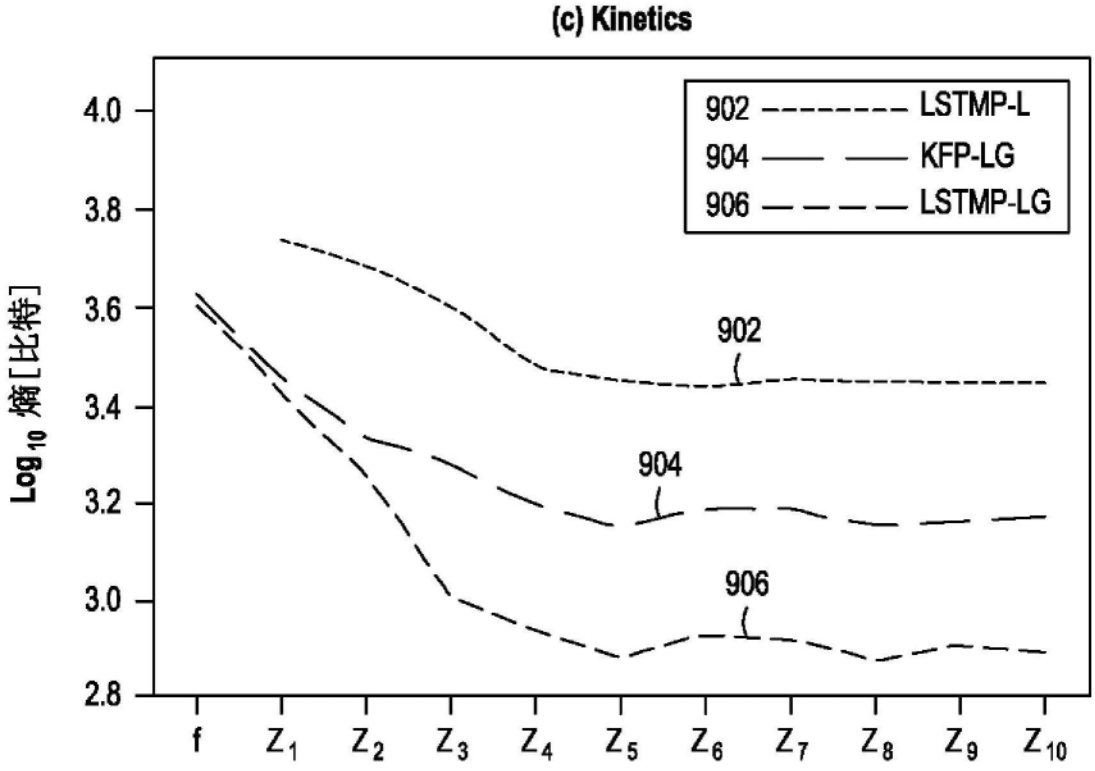

(c) Kinetics
4.0
3.8
3.6
3.4
3.2
3.0
2.8
Log10 熵[比特]
902
904
906
902 LSTMP-L
904 KFP-LG
906 LSTMP-LG
f
Z1
Z2
Z3
Z4
Z5
Z6
Z7
Z8
Z9
Z10

图9C

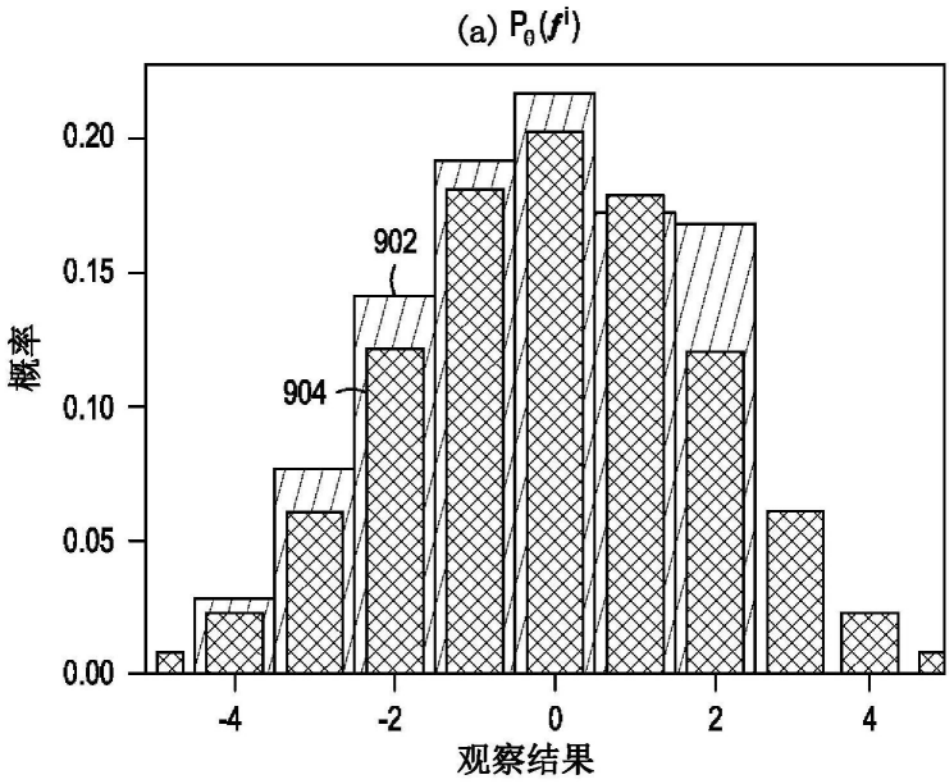

(a) $P_\theta(f^i)$
0.20
0.15
0.10
0.05
0.00
902
904
概率
-4
-2
0
2
4
观察结果

图10A

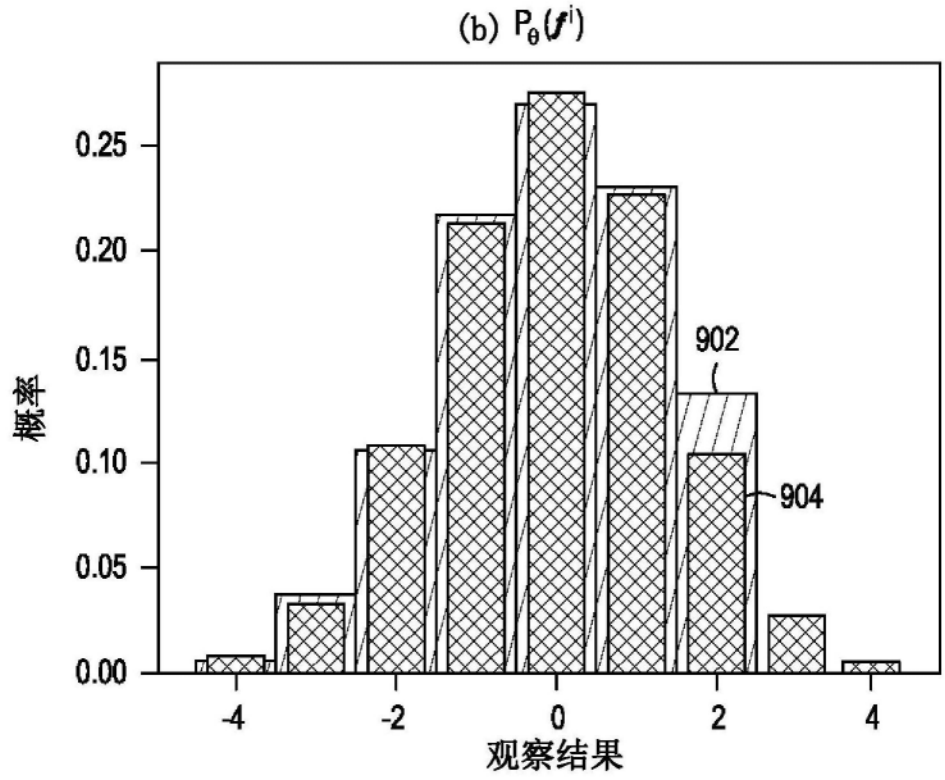

(b) $P_\theta(f^i)$
0.25
0.20
0.15
0.10
0.05
0.00
902
904
概率
-4
-2
0
2
4
观察结果

图10B

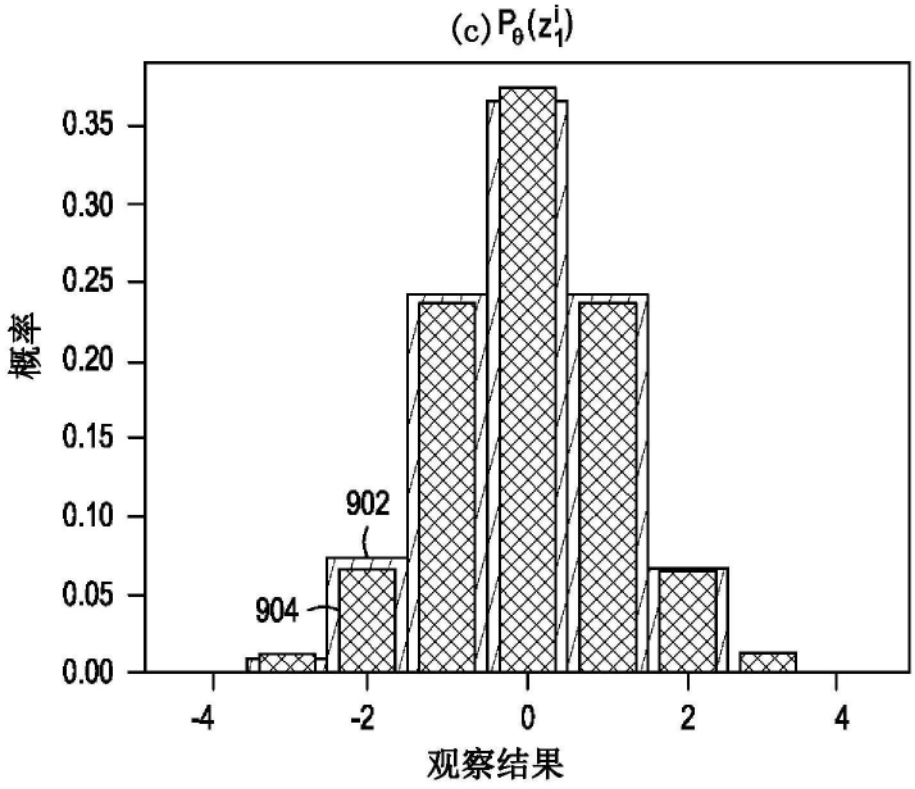

(c) $P_\theta(z_1^i)$
概率
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
902
904
-4
-2
0
2
4
观察结果

图10C

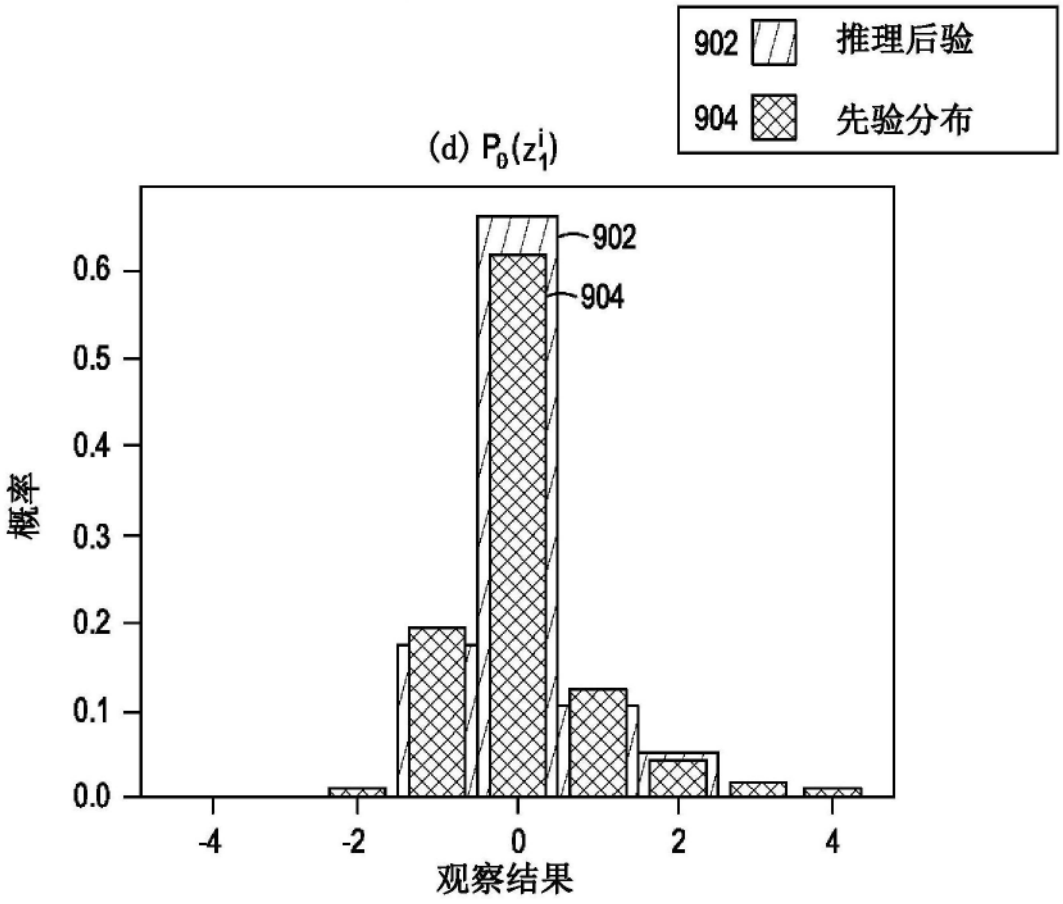

902 推理后验
904 先验分布
(d) $P_0(z_1^i)$
902
904
概率
0.0
0.1
0.2
0.3
0.4
0.5
0.6
-4
-2
0
2
4
观察结果

图10D

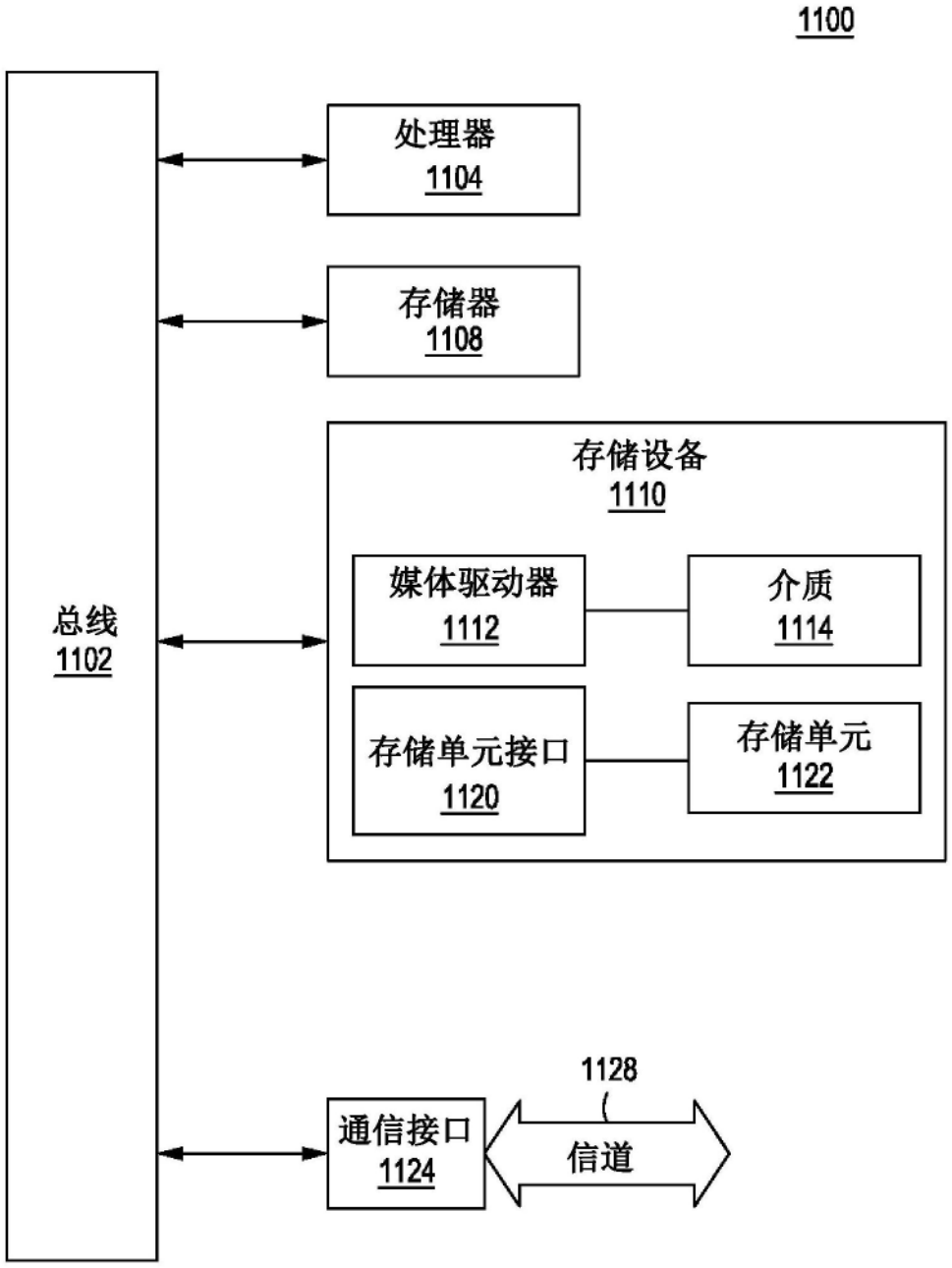

1100
处理器
1104
存储器
1108
存储设备
1110
媒体驱动器
1112
介质
1114
存储单元接口
1120
存储单元
1122
总线
1102
1128
通信接口
1124
信道

图11