



(12) 发明专利

(10) 授权公告号 CN 111465979 B

(45) 授权公告日 2025. 03. 07

(21) 申请号 201980006424.X

(22) 申请日 2019.10.10

(65) 同一申请的已公布的文献号
申请公布号 CN 111465979 A

(43) 申请公布日 2020.07.28

(30) 优先权数据
2018-197972 2018.10.19 JP
2019-061040 2019.03.27 JP

(85) PCT国际申请进入国家阶段日
2020.06.12

(86) PCT国际申请的申请数据
PCT/JP2019/039939 2019.10.10

(87) PCT国际申请的公布数据
W02020/080239 JA 2020.04.23

(73) 专利权人 索尼公司
地址 日本东京

(72) 发明人 赤间健人

(74) 专利代理机构 北京康信知识产权代理有限公司 11240
专利代理师 余刚

(51) Int.Cl.
G10H 1/00 (2006.01)
G06N 3/02 (2006.01)
G06N 20/00 (2006.01)

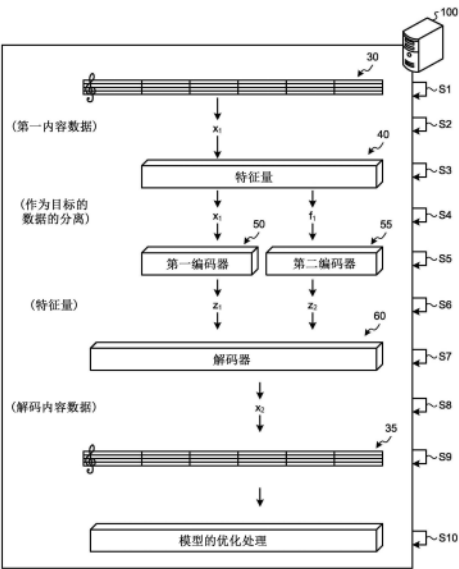
(56) 对比文件
GINO BRUNNER et,al..《MIDI-VAE:
Modeling Dynamics and Instrumentation of
Music with Application to Style
Transfer》.《ARXIV.ORG,CORNELL UNIVERSITY
LIBRARY》.2018,1-8.

审查员 张陈磊

权利要求书2页 说明书23页 附图18页

(54) 发明名称
信息处理方法、信息处理装置和信息处理程序

(57) 摘要
本信息处理装置 (100) 包括：提取单元 (131)，从构成第一内容的元素中提取第一数据；和模型生成单元 (132)，生成具有第一编码器 (50) 和第二编码器 (55) 的训练模型，第一编码器计算作为第一内容的特征量的第一特征量，第二编码器计算作为所提取的第一数据的特征量的第二特征量。



1. 一种信息处理装置,包括:

提取单元,被配置为:

接收由用户为第一内容指定的提取操作和执行所述提取操作的范围的指定;

通过对接收的范围上的数字数据执行提取操作,从构成作为歌曲的所述第一内容的数字数据中提取第一数据,其中,所述数字数据表示歌曲的声音信息并且指示音高、声长和休止符;和

模型生成单元,被配置为生成具有第一编码器、第二编码器和解码器的学习模型,

其中,所述第一编码器被训练为从所述数字数据提取作为所述第一内容的特征量的第一特征量,

其中,所述第二编码器被训练为从所述第一数据提取作为所提取的第一数据的特征量的第二特征量,并且

其中,所述解码器被训练为基于所述第一特征量和所述第二特征量来解码与所述第一内容相对应的数据;

其中,生成所述学习模型包括:

将所述数字数据输入至所述第一编码器,

将所述第一数据输入至所述第二编码器,

将由所述第一编码器输出的所述第一特征量和由所述第二编码器输出的所述第二特征量输入至所述解码器,

基于所述解码器输出的数据重构所述第一内容,以及

调整所述第一编码器、所述第二编码器和所述解码器的参数,以便减小所述第一内容与所重构的第一内容之间的重构误差。

2. 根据权利要求1所述的信息处理装置,其中

所述提取单元,被配置为

通过对所述范围内包括的所述数字数据执行具有顺序不变性的运算处理作为所述提取操作来提取所述第一数据。

3. 根据权利要求1所述的信息处理装置,其中

所述提取单元,被配置为

从所述范围内包括的所述数字数据中,通过丢弃与时间无关或与顺序无关的信息来提取所述第一数据。

4. 根据权利要求1所述的信息处理装置,其中

所述提取单元,被配置为

执行丢失与所述范围内包括的所述声音信息中的声音的音高有关的信息的提取操作并且从所述第一内容中提取所述第一数据。

5. 根据权利要求1所述的信息处理装置,进一步包括:

内容生成单元,被配置为通过向所述解码器输入从经学习的第一编码器或第二编码器输出的特征量或组合基于所述特征量生成的特征量中的任意一个而获得的值来生成第二内容,所述第二内容是与所述第一内容具有相同格式的新内容。

6. 根据权利要求5所述的信息处理装置,其中

所述内容生成单元,被配置为

通过向所述解码器输入将从经学习的所述第一编码器输出的特征量与从经学习的所述第二编码器输出的特征量进行组合而获得的值来生成所述第二内容。

7. 根据权利要求5所述的信息处理装置, 其中

所述内容生成单元, 被配置为

通过固定所组合的值之中的与任意一个特征量对应的值并且按顺序改变与其他特征量对应的值, 来按顺序生成多个所述第二内容。

8. 一种信息处理方法, 包括:

允许计算机:

接收由用户为第一内容指定的提取操作和执行所述提取操作的范围的指定;

通过对接收的范围上的数字数据执行提取操作, 从构成作为歌曲的所述第一内容的数字数据中提取第一数据, 其中, 所述数字数据表示歌曲的声音信息并且指示音高、声长和休止符; 以及

生成具有第一编码器、第二编码器和解码器的学习模型,

其中, 所述第一编码器被训练为从所述数字数据提取作为所述第一内容的特征量的第一特征量,

其中, 所述第二编码器被训练为从所述第一数据提取作为所提取的第一数据的特征量的第二特征量, 并且

其中, 所述解码器被训练为基于所述第一特征量和所述第二特征量来解码与所述第一内容相对应的数据;

其中, 生成所述学习模型包括:

将所述数字数据输入至所述第一编码器,

将所述第一数据输入至所述第二编码器,

将由所述第一编码器输出的所述第一特征量和由所述第二编码器输出的所述第二特征量输入至所述解码器,

基于所述解码器输出的数据重构所述第一内容, 以及

调整所述第一编码器、所述第二编码器和所述解码器的参数, 以便减小所述第一内容与所重构的第一内容之间的重构误差。

9. 一种计算机可读存储介质, 其上存储有包括指令的程序, 当被计算机执行时, 使得所述计算机执行根据权利要求8所述的方法。

信息处理方法、信息处理装置和信息处理程序

技术领域

[0001] 本公开涉及一种信息处理方法、信息处理装置以及信息处理程序。更具体地,本公开涉及一种通过机器学习生成的学习模型的生成处理。

背景技术

[0002] 在各种技术领域中,已使用了利用机器学习的信息处理。例如,通过使用模拟大脑神经系统的机制的神经网络学习内容(图像、音乐等)的特征,自动地生成其他内容。

[0003] 例如,已经提出了一种技术,该技术使用户能够通过学习现有歌曲的特征来自动创作与歌词匹配的适当歌曲,而无需用户输入除歌词以外的参数。

[0004] 现有技术文献

[0005] 专利文献

[0006] 专利文献1:特开2011-175006号公报

发明内容

[0007] 技术问题

[0008] 根据相关技术,通过学习根据表示每首歌曲的歌词的歌词数据和表示歌曲的属性的属性数据计算出的语言特征量,如果给出新的歌词数据,则可以自动生成与新的歌词数据相匹配的歌曲。

[0009] 然而,在相关技术中,仅生成与歌词匹配的旋律或和弦,并不总是以用户所希望的方式生成歌曲。例如,在相关技术中,难以生成保留用户所指定的特征的歌曲,诸如,在仅保留特定的现有歌曲的节奏的同时改变旋律等。也就是,在传统的处理中,难以执行选择性地学习内容的某些特征(例如,歌曲的节奏、旋律及构成声音)的处理。

[0010] 因此,本公开提出了一种能够选择性地学习用户所指定的特征量的信息处理装置、信息处理方法以及信息处理程序。

[0011] 问题的解决方案

[0012] 为了解决上述问题,一种信息处理装置包括:提取单元,从构成第一内容的元素中提取第一数据;和模型生成单元,生成具有第一编码器和第二编码器的学习模型,第一编码器计算作为第一内容的特征量的第一特征量,第二编码器计算作为所提取的第一数据的特征量的第二特征量。

附图说明

[0013] 图1是示出根据本公开的实施方式的信息处理的示例的示意图。

[0014] 图2是示出根据本公开的实施方案的特征量的示意图。

[0015] 图3是用于对根据本公开的实施方案的生成处理进行说明的示意图。

[0016] 图4是示出根据本公开的实施方案的信息处理装置的配置例的示意图。

[0017] 图5是示出根据本公开的实施方案的歌曲数据存储单元的示例的示意图。

- [0018] 图6是示出根据本公开的实施方式的信息处理步骤的示例的流程图(1)。
- [0019] 图7是示出根据本公开的实施方式的信息处理步骤的流程图(2)。
- [0020] 图8是示出根据本公开的第一变形的学习处理的示例的示意图。
- [0021] 图9是示出根据本公开的第二变形的学习处理的示例的示意图。
- [0022] 图10是示出根据本公开的第三变形的生成处理的示例的示意图。
- [0023] 图11是用于对根据本公开的生成处理进行概念性地描述的示意图。
- [0024] 图12是示出根据本公开的第四变形的学习处理的示例的示意图(1)。
- [0025] 图13是示出根据本公开的第四变形的学习处理的示例的示意图(2)。
- [0026] 图14是示出根据本公开的第四变形的学习处理的示例的示意图(3)。
- [0027] 图15是示出根据本公开的第四变形的学习处理的示例的示意图(4)。
- [0028] 图16是示出根据本公开的第五变形的学习处理的示例的示意图。
- [0029] 图17是示出根据本公开的第五变形的生成处理的示例的示意图(1)。
- [0030] 图18是示出根据本公开的第五变形的生成处理的示例的示意图(2)。
- [0031] 图19是示出根据本公开的第六变形的信息处理的示例的示意图(1)。
- [0032] 图20是示出根据本公开的第六变形的信息处理的示例的示意图(2)。
- [0033] 图21是示出根据本公开的第六变形的信息处理的示例的示意图(3)。
- [0034] 图22是示出实现信息处理装置的功能的计算机的示例的硬件配置图。

具体实施方式

[0035] 下面将参考附图对本公开的实施方案进行详细描述。在下列各个实施方式中,以相同参考标号表示相同的零件,并且将省去其重复性的描述。

[0036] 将按照下列各项的顺序对本公开进行描述。

[0037] 1.实施方式

[0038] 1-1.根据本公开的实施方案的信息处理的概要

[0039] 1-2.根据实施方式的信息处理装置的配置

[0040] 1-3.根据实施方式的信息处理步骤

[0041] 2.其他实施方式

[0042] 3.硬件配置

[0043] (1.实施方式)

[0044] [1-1.根据本公开的实施方案的信息处理的概要]

[0045] 图1是示出根据本公开的实施方案的信息处理的示例的示意图。通过图1中示出的信息处理装置100实现根据本公开的实施方案的信息处理。

[0046] 信息处理装置100是执行根据本公开的信息处理的装置并且是例如信息处理终端或服务器装置。

[0047] 在实施方式中,信息处理装置100生成用于提取内容的特征的学习模型并且基于所生成的学习模型来生成新的内容。在实施方式中,内容由诸如音乐(歌曲)、图像以及动画(运动图像)等预定格式的数字数据构成。在图1的示例中,使用信息处理装置100对作为内容的示例的歌曲进行处理。

[0048] 根据实施方式的学习模型具有编码器和解码器,编码器从构成内容的数据中提取

特征量,并且解码器根据所提取的特征量对内容进行重构。例如,信息处理装置100通过诸如变分自编码器(VAE)和生成对抗网络(GAN)等无监督学习来学习编码器。具体地,信息处理装置100将内容输入至编码器,从所提取的特征量对内容进行重构,对重构内容之前和之后的内容进行比较并且对编码器和解码器的参数进行调整。信息处理装置100通过重复该处理并且优化编码器和解码器的参数而生成学习模型。因此,信息处理装置100能够生成从构成内容的数据中获得适当特征量的学习模型。应注意,特征量由例如具有的维数比输入内容的数据更低的矢量表示。

[0049] 应注意,学习模型并不局限于上述示例,并且只要学习模型能够提取特征量并且从所提取的特征量对内容进行重构,则学习模型可以是任意模型。

[0050] 在图1的示例中,信息处理装置100将歌曲(换言之,表示构成歌曲的声音的数字数据)输入至编码器,并且提取诸如旋律线、构成声音、节奏(音乐的时间结构,例如,包括多少音符和休止符,按什么顺序再生声音等)、速度以及节拍等表示歌曲的特征的元素,作为歌曲的特征量。

[0051] 此处,在如上所述生成用于提取现有歌曲的特征量的学习模型的情况下,适当地提取用户所指定的特征可能是一个问题。例如,如果按原样提取任意歌曲的特征量,则可以再现歌曲的特征,但是有时难以使用诸如使用特征量生成新内容的特征量。具体地,使用上述VAE等提取的特征量是一列矢量数值,人类难以识别每个数值表示什么特征。例如,即使自动生成仅包含现有歌曲的节奏特征的新歌曲,人类也难以判断是否使用任意数值生成特征量矢量(换言之,表示特征量的潜在空间)。

[0052] 因此,根据本公开的信息处理装置100通过下述信息处理生成不仅能够提取整首歌曲的特征量,而且还能够提取用户所希望的特征量的学习模型。具体地,信息处理装置100从作为计算特征量所针对的目标的内容(在下文中,为了区分被称为“第一内容”)中包含的元素(换言之,构成第一内容的数字数据)中提取预定数据(下文中,为了区分被称为“第一数据”)。信息处理装置100生成具有第一编码器和第二编码器的学习模型,第一编码器计算作为第一内容的特征量的第一特征量,第二编码器计算作为所提取的第一数据的特征量的第二特征量。

[0053] 以这种方式,信息处理装置100生成具有包括第一编码器和第二编码器的配置的学习模型,当提取第一内容的特征量时,第一编码器提取全部第一内容的特征量,第二编码器仅处理第一内容的特定数据。尽管在后面进行了详细描述,然而,第一数据是通过预定操作从第一内容中提取的数据,以易于提取用户所指定的特征量。以这种方式,信息处理装置100生成能够通过使用多个编码器从第一内容及第一内容的某些元素中分别提取特征量的学习模型。结果,信息处理装置100能够获得其中仅用户所希望的特征被单独表达的特征量。此外,信息处理设备100可以使用单独获得的特征量而仅改变或不改变用户在现有内容中指定的特征(例如,歌曲的节奏),从而使得可以自动地生成新的内容(以下,为了区别而称为“第二内容”)。为了说明的目的,将提取第一内容的特征量的编码器称为第一编码器,将提取第一数据的特征量的编码器称为第二编码器,但这是为了方便。例如,提取第一内容的特征量的编码器可以是第二编码器。

[0054] 在下文中,参考图1,将根据流程描述本公开的信息处理的概要。图1示出了其中信息处理装置100使用学习数据来使学习模型(第一编码器50、第二编码器55以及解码器60)

进行学习的处理的流程。

[0055] 首先,信息处理装置100获取作为学习数据的示例的歌曲30(步骤S1)。歌曲30由例如表示音高、音长以及休止符的符号串(数字数据)构成。作为示例,音高是表示以预定阶数(例如,128阶等)表示声音的音高的频率的音高。此外,音长表示再现的声音保持多长时间。此外,休止符表示声音再现停止的时刻。进一步地,指示歌曲30的数据可以包括诸如歌曲30的节拍或速度、指示小节的停顿的符号、特定时刻的和弦、或构成和弦的构成声音等信息。

[0056] 例如,信息处理装置100对符号串进行转换,以使得其能够通过模型进行处理。作为示例,信息处理装置100使用被表示成矢量(例如,表示音高、音长等分配给每个维度的声音信息的嵌入(embedding)矢量)的符号串进行处理。表示该声音的数据可以由例如乐器数字接口(MIDI)(注册商标)格式表示,可以是由通用音序器再现的已知格式的数字数据,或可以被表示成WAV格式等波形数据。

[0057] 通过上述处理,信息处理装置100获得表示歌曲30的数据(步骤S2)。在图1的示例中,表示歌曲30的数据(矢量)被表达成“ x_1 ”。

[0058] 接着,信息处理装置100对数据 x_1 执行提取操作40(步骤S3)。提取操作40是仅从歌曲30中分离出特定特征(例如,用户所希望的特征)的预处理,意指对数据 x_1 执行预定操作并且提取预定数据(对应“第一数据”)的操作。

[0059] 作为示例,信息处理装置100通过对数据 x_1 执行具有顺序不变性的运算处理来提取第一数据(步骤S4)。例如,具有顺序不变性的运算是使用交换运算符的运算。具体地,具有顺序不变性的运算是指不取决于顺序的运算(例如,包括可指具有近似顺序不变性的运算,使得计算结果通过包括预定系数而略微地改变),诸如加法或乘法以及取最大值(max)运算等。应注意,在图1的示例中,将通过提取操作40从数据 x_1 中提取的信息表示成特征信息“ f_1 ”。

[0060] 例如,在图1的示例中,假设使信息处理装置100执行处理的用户希望从歌曲30的特征中提取“节奏”。歌曲中的节奏是与时间(顺序)有关的特征。为此,例如,当试图从歌曲30中仅提取节奏时,可以考虑留下表示音符的布置(音长与休止符之间的关系)的信息及隐藏(掩盖)其他信息的方法。

[0061] 例如,信息处理装置100对数据 x_1 中包括的数据执行具有顺序不变性的运算并且丢弃“表示声音的顺序的信息”。作为示例,信息处理装置100能够将数据 x_1 中的嵌入矢量的平均值作为具有顺序不变性的运算。

[0062] 将基于特定示例对该方法进行描述。假设与构成歌曲30的8个节拍对应的音符的布置是“C4,_,_,_,E4,_,_,R”。在本示例中,“C4”和“E4”表示声音的音高。此外,符号“_”表示对之前的声音进行延长。此外,“R”表示“休止符”。

[0063] 此处,假设信息处理装置100具有表示与每个声音或休止符对应的嵌入矢量的数据表。例如,嵌入矢量是d维度(d是任意整数)并且是其中“1”输入至与相应声音对应的维度(诸如,“C4”或“E4”)并且“0”输入至其他维度的矢量。

[0064] 即,当参考诸如“C4,_,_,_,E4,_,_,R”等数据的布置时,信息处理装置100参考“C4”的嵌入矢量四次,然后参考“E4”的嵌入矢量三次,以及参考“R”的嵌入矢量一次。

[0065] 此处,信息处理装置100对与“C4,_,_,_,E4,_,_,R”对应的嵌入矢量求平均(可以是加法)。然后,在平均矢量中,丢弃了取决于顺序的信息(即,节奏),并且提取除表示声音

的顺序的信息之外的信息(例如,构成声音等)。以这种方式,信息处理装置100能够提取歌曲的局部部分中的“除表示声音的顺序的信息之外的信息”。

[0066] 在通过这种方式获得的特征信息 f_1 中,保留数据 x_1 中所包括的构成声音(即,构成歌曲30的构成声音的音高),但是,丢弃表示节奏的信息(发出声音的时刻、休止符所处的时刻等)。

[0067] 随后,信息处理装置100将原始数据 x_1 输入至第一编码器50并且将特征信息 f_1 输入至第二编码器55,第二编码器55是与第一编码器50不同的编码器(步骤S5)。应注意,例如,第一编码器50和第二编码器55是上述VAE等中的编码器并且提取输入数据的特征量作为低维矢量。

[0068] 信息处理装置100获得从第一编码器50和第二编码器55中的每一个输出的特征量(步骤S6)。在图1的示例中,将数据 x_1 的特征量表达成“ z_1 ”,并且将特征信息 f_1 的特征量表达成“ z_2 ”。

[0069] 应注意,在描述中对特征量 z_1 和特征量 z_2 进行了单独描述,但是,特征量 z_1 和特征量 z_2 实际上被包括在一个特征量矢量中。将参考图2对此进行描述。图2是用于对根据本公开的実施方式的特征量进行说明的示意图。

[0070] 如图2中示出的,假设在通过信息处理装置100生成的学习模型中,特征量矢量是64维矢量。在这种情况下,从第一编码器50输出的特征量 z_1 构成特征量矢量的一部分(例如,特征量矢量的前半部分)。此外,从第二编码器55输出的特征量 z_2 也构成特征量矢量的一部分(例如,特征量矢量的后半部分)。应注意,将特征量矢量设置成64维,将多少维度分配给第一编码器50的输出,将多少维度分配给第二编码器55的输出等,可任意设置。

[0071] 也就是,通过使用多个编码器构成模型,信息处理装置100能够判断特征量矢量的某个维度是从某个编码器输出的值。即,信息处理装置100能够输出能人工解释某个维度表示某个特征的特征量矢量。

[0072] 返回至图1,继续进行描述。信息处理装置100将从第一编码器50和第二编码器55中的每一个输出的特征量 z_1 和特征量 z_2 输入至解码器60(步骤S7)。

[0073] 解码器60是经过学习基于通过编码器提取的特征量重构内容的解码器。在图1的示例中,解码器60基于从第一编码器50和第二编码器55中的每一个输出的特征量 z_1 和特征量 z_2 来输出数据 x_2 (步骤S8)。

[0074] 数据 x_2 与作为第一内容的数据的数据 x_1 具有相同的格式。即,数据 x_2 可以指用于再现与歌曲30具有相同格式的歌曲35的数据(符号串)。换言之,信息处理装置100基于从解码器60输出的数据 x_2 来重构内容(在图1的示例中为歌曲35)(步骤S9)。

[0075] 然后,信息处理装置100基于重构内容执行模型的学习处理。具体地,信息处理装置100对特征量的提取处理之前和之后的数据(即,数据 x_1 和数据 x_2)进行比较,并且优化第一编码器50、第二编码器55以及解码器60的参数使得数据 x_2 接近数据 x_1 (步骤S10)。更具体地,信息处理装置100对参数进行优化,以使得数据 x_1 与数据 x_2 之间的重构误差减小并且特征量的后验分布接近先验分布。进一步地,信息处理装置100获取除数据 x_1 之外的多条学习数据并且重复图1中示出的学习处理,以优化学习模型。因此,信息处理装置100能够根据本公开生成学习模型。

[0076] 根据通过图1的处理生成的学习模型,当将新的歌曲输入至模型时,可以获得其中

分离出歌曲的特征量的特征量矢量。在图1的示例中,因为根据学习模型的第二编码器55基于学习数据之中的其中已经丢弃节奏信息的数据(例如,特征信息 f_1)进行学习,所以第二编码器55是经过学习提取除节奏之外的特征的编码器。换言之,从第二编码器55输出的特征量可以指捕获诸如歌曲中的构成声音等顺序不变特征量的一个特征量(可由人类解释为某个维度指示“构成声音”的一个特征量)。另一方面,根据学习模型的第一编码器50可以指经过学习从特征量矢量中主要提取顺序可变特征量(即,节奏)的编码器。换言之,从第一编码器50输出的特征量可以指仅捕获歌曲中的节奏特征的一个特征量(可由人类解释为某个维度指示“节奏”的一个特征量)。

[0077] 如上所述,信息处理装置100将所分离和提取的信息提前输入至多个编码器并且基于信息执行学习,由此使得可以将通常表示为简单的数值的序列的特征量矢量分离成特定的特征。例如,如果内容是歌曲,信息处理装置100则能够输出其中分离出节奏或旋律(诸如和弦的构成声音等)的特征量矢量。即,信息处理装置100能够生成分离用户所指定的特征(能够选择性地学习用户所指定的特征量)的学习模型。

[0078] 此外,信息处理装置100还能够通过使用所生成的学习模型而生成具有用户所指定的特征的新内容(第二内容)。将参考图3对此进行描述。图3是用于对根据本公开的实施方式的生成处理进行说明的示意图。

[0079] 如图3中示出的,信息处理装置100获取图1中示出的特征量 z_1 (步骤S11)。如上所述,特征量 z_1 是包括相对更多的歌曲30中的与节奏有关的特征的特征量。换言之,特征量 z_1 是从歌曲30中分离出的针对目标的特征量。

[0080] 另一方面,信息处理装置100获取特征量 z_3 ,其是与图1中示出的特征量 z_2 具有相同维度的特征量(步骤S11)。特征量 z_3 表示任意数据的特征量。例如,任意数据的特征量是使用第二编码器55从歌曲30之外的其他歌曲中提取的特征量。可替代地,任意数据的特征量还可以是从诸如特征量 z_2 等特征量的先验分布中采样的特征量。

[0081] 然后,信息处理装置100将特征量 z_1 和特征量 z_2 输入至解码器60(步骤S12)。信息处理装置100获取从解码器60输出的数据 x_3 (步骤S13)。随后,信息处理装置100从数据 x_3 重构歌曲45作为第二内容(步骤S14)。

[0082] 如上所述,特征量 z_1 是表示歌曲30的节奏特征的特征量。出于此原因,由数据 x_3 重构的歌曲45具备歌曲30的节奏特征(例如,音符的布置顺序、音符的划分、小节的数量、速度等),并且构成声音等是具有由特征量 z_3 表示的特征的歌曲。具体地,歌曲45在保留歌曲30的节奏的同时,具有的由该节奏演奏的音阶(例如,反映特征量 z_3 中的构成声音的特征的音阶)不同于歌曲30的音阶的歌曲。应注意,在重构歌曲45之后,信息处理装置100可以在固定特征量 z_1 (固定节奏)的同时重复从先验分布中对特征量 z_3 的重新采样。结果,信息处理装置100能够自动生成在保留相同节奏特征的同时以不同的音阶播放的多首歌曲。此处,在生成第二内容时,特征量 z_1 不必与图1中示出的特征量 z_1 相同,并且可以使用从特征量 z_1 的先验分布中采样的值。

[0083] 由此,信息处理装置100能够分离并且提取特定的特征(在本示例中,为节奏)并且使用所提取的特征量生成内容,结果,能够自由创作具有用户所希望的特征的歌曲。例如,信息处理装置100能够在固定节奏的同时生成具有不同构成声音的歌曲的变体。此外,当不固定节奏而是固定构成声音的特征(在上述示例中,固定特征量 z_2 ,并且将不同的特征量输

入至与特征量 z_1 对应的位置)时,信息处理装置100能够生成具有固定的和弦类型的各种瑟音变化,诸如播放各种节奏(音相)等。换言之,如果特征是由用户能够指定的特征(换言之,能够描述生成特征信息的规则),则信息处理装置100固定任意音乐概念从而能够自动生成具有各种变化的歌曲。

[0084] 应注意,在上述步骤S4中,已经示出了其中信息处理装置100使用数据 x_1 中的顺序不变性提取关于节奏的特征信息 f_1 的示例。此处,信息处理装置100可以通过使用丢弃表示音高的信息的方法来提取特征信息 f_1 。

[0085] 基于与上述步骤S4中的示例类似的特定示例给出描述。假设与构成歌曲30的8个节拍对应的音符的布置是“C4,_,_,_,E4,_,_,R”。在本示例中,“C4”和“E4”表示声音的音高。此外,符号“_”表示对之前的声音进行延长。此外,“R”表示“休止符”。

[0086] 此处,信息处理装置100删除表示音高的信息。例如,信息处理装置100以诸如“M”等公共信息替代诸如“C4”和“E4”等信息。结果,构成歌曲30的信息是“M,_,_,_,M,_,_,R”。该信息是其中仅丢弃音高的信息,但该信息中保留了节奏。

[0087] 在这种情况下,因为所提取的信息(被称为“特征信息 f_2 ”)不同,所以在步骤S5之后编码器的学习也不同。即,因为特征信息 f_2 具有保留歌曲30的节奏的信息,所以学习该信息的第二编码器55经学习输出表示歌曲的“节奏特征”的特征量 z_2 。另一方面,第一编码器50经学习输出歌曲的除“节奏特征”之外的特征量 z_1 。即,训练(学习)了提取与图1中示出的示例不同的信息的编码器。以这种方式,第一编码器50和第二编码器55的学习预先根据提取操作40的类型而变化。

[0088] 应注意,图1中的示例示出了学习整首歌曲30的特征的示例,但是,信息处理装置100可以在每个预定的部分暂停歌曲30并且获取每部分中的内容的数据(部分数据)。

[0089] 在这种情况下,信息处理装置100可以使用各种已知的方法将歌曲30划分成部分数据。例如,信息处理装置100检测小节的预设停顿并且将歌曲30划分成部分数据(小节)。可替代地,信息处理装置100可以基于节拍与音符之间的关系将歌曲30划分成部分数据。例如,如果歌曲30是4/4节拍,信息处理装置100则检测具有与四分音符对应的的长度的四个声音再现时的时间点作为一个停顿并且将歌曲30划分成小节。

[0090] 此外,信息处理装置100可以使用除小节之外的停顿(pause)作为部分数据。例如,信息处理装置100可以检测歌曲30的旋律线的停顿(例如,具有超过预定阈值的长度的休止符出现的位置等)并且将歌曲划分成部分数据。在这种情况下,部分数据不一定需要与小节相匹配。进一步地,信息处理装置100可以对歌曲30的和弦改变时的时刻等进行检测并且可以基于所检测的时刻将歌曲30划分成部分数据。

[0091] [1-2. 根据实施方式的信息处理装置的配置]

[0092] 接着,将描述作为根据实施方式的执行信息处理的信息处理装置的示例的信息处理装置100的配置。图4是示出根据本公开的实施方式的信息处理装置100的配置例的示图。

[0093] 如图4中示出的,信息处理装置100具有通信单元110、存储单元120以及控制单元130。应注意,信息处理装置100可以具有从管理信息处理装置100的管理员等接收各种操作的输入单元(例如,键盘或鼠标)或显示各种信息的显示单元(例如,液晶显示器等)。

[0094] 例如,通过网络接口卡(NIC)等实现通信单元110。通信单元110通过有线或无线方式连接至网络N(诸如互联网等)并且经由网络N将信息发送至其他装置等和从其他装置等

接收信息。

[0095] 例如,通过诸如随机存取存储器(RAM)或闪存等半导体存储器元件、或诸如硬盘或光盘等存储设备实现存储单元120。存储单元120包括模型存储单元121和歌曲数据存储单元122。

[0096] 模型存储单元121存储学习模型。具体地,模型存储单元121包括提取内容的特征量的第一编码器50和第二编码器55以及重构内容的解码器60。应注意,模型存储单元121可以存储诸如学习所使用的内容等学习数据。

[0097] 歌曲数据存储单元122存储与被输入至模型的内容(歌曲)有关的数据。图5示出了根据实施方式的歌曲数据存储单元122的示例。图5是示出根据本公开的实施方式的歌曲数据存储单元122的示例的示图。在图5示出的示例中,歌曲数据存储单元122具有诸如“歌曲ID”、“部分数据ID”、“音高信息”、“音长休止符信息”、“和弦信息”以及“节奏信息”等项。

[0098] “歌曲ID”是用于识别歌曲的识别信息。“部分数据ID”是用于识别部分数据的识别信息。例如,部分数据与构成歌曲的一个或多个小节对应。

[0099] “音高信息”表示关于部分数据中所包括的声音的音高(音阶)的信息。“音长休止符信息”表示部分数据中所包括的声音的长度(再现时间或再现节拍)或休止符的长度或时刻。“和弦信息”表示部分数据中所包括的和弦的类型、和弦的构成声音、小节中的和弦的切换等。“节奏信息”表示小节的节拍或速度(tempo)、强节拍的位置、弱节拍的位置等。

[0100] 应注意,在图5中,将诸如音高信息等项概念性地描述为“C01”,但实际上,每项存储表示上述声音的特定数据。此外,出于说明之缘故,图5示出了其中将“音高信息”、“音长休止符信息”等存储为不同的项的示例,但是这些条信息可以作为指示小节中所包括的音符的信息而被统一存储在一项等中。即,表示歌曲的数据的格式并不局限于图5中示出的格式,而是只要格式能够被模型处理,可以是任意格式。

[0101] 例如,图5中示出的示例示出了通过歌曲ID“A01”识别的歌曲具有通过部分数据ID“B01”或“B02”识别的部分数据。此外,具有部分数据ID“B01”的部分数据包括其音高信息是“C01”、音长休止符信息是“D01”、和弦信息是“E01”及节奏信息是“F01”的声音数据。

[0102] 返回至图4,将继续进行描述。通过以随机存取存储器(RAM)等作为工作区,使信息处理装置100中所存储的程序(例如,根据本公开的信息处理程序)运行的例如中央处理单元(CPU)、微处理单元(MPU)等来实现控制单元130。此外,控制单元130是控制器并且可以通过诸如专用集成电路(ASIC)或现场可编程门阵列(FPGA)等集成电路来实现。

[0103] 如图4中示出的,控制单元130包括提取单元131、模型生成单元132、获取单元133以及内容生成单元134,并且实现或执行下述信息处理的功能或操作。应注意,控制单元130的内部配置并不局限于图4中示出的配置,而是只要控制单元130执行后面所述的信息处理,则可以是其他配置。

[0104] 提取单元131从构成第一内容的元素中提取第一数据。例如,提取单元131根据用户的指定,基于预定规则从构成第一内容的元素中提取第一数据。例如,提取单元131对构成作为歌曲的第一内容的数字数据执行提取操作,以提取第一数据。

[0105] 例如,提取单元131接收用户针对第一内容而指定的操作及执行操作的范围的指定,并且在所接收的范围内执行操作。具体地,在图1示出的歌曲30中,提取单元131从用户接收仅从歌曲30中提取节奏的操作及提取节奏的范围(例如,小节等)的指定。然后,提取单

元131根据用户的指定执行与提取有关的操作。

[0106] 作为操作示例,提取单元131通过对范围中所包括的元素执行具有顺序不变性的运算处理来提取第一数据。例如,提取单元131对构成第一内容的数据执行诸如加法、乘法、或取最大值运算等的顺序不变运算。通过该处理,提取单元131能够从第一内容中提取其中仅与顺序有关的信息(在歌曲的示例中为节奏)被掩盖的第一数据。

[0107] 应注意,从范围内所包括的元素中,提取单元131可以通过丢弃不是时间相关或顺序相关的信息来提取第一数据。在歌曲中,不是时间相关或顺序相关的信息包括例如特定小节中的和弦的构成声音、歌曲的调性、音符的总数、歌曲的节拍等。另一方面,与时间或顺序有关的信息的示例包括歌曲的节奏信息。

[0108] 例如,提取单元131能够通过丢弃第一内容的特定范围的信息之中作为与时间无关或与顺序无关的信息的音高信息,来提取其中仅保留了节奏信息的第一数据。

[0109] 模型生成单元132生成具有第一编码器50和第二编码器55的学习模型,第一编码器50计算作为第一内容的特征量的第一特征量,第二编码器55计算作为提取单元131所提取的第一数据的特征量的第二特征量。应注意,学习模型中所包括的编码器的数量并不局限于两个,并且学习模型中可以包括更多数量的编码器。

[0110] 进一步地,模型生成单元132生成包括解码器60的学习模型,解码器60基于第一特征量和第二特征量解码与第一内容对应的数据。即,当对从多个编码器中提取的特征量进行解码时,模型生成单元132生成被配置为通过一个共同解码器60进行解码的学习模型。

[0111] 例如,在学习诸如歌曲等内容数据的情况下,模型生成单元132将内容数据输入至第一编码器50并且将从内容数据中提取的特征信息输入至第二编码器55。随后,模型生成单元132将从第一编码器50和第二编码器55输出的特征量输入至解码器60并且重构内容(歌曲)。然后,模型生成单元132对第一编码器50、第二编码器55以及解码器60的参数进行调整,以使得之前内容之后内容的数据之差减少。模型生成单元132重复该处理并且生成其中对第一编码器50、第二编码器55、以及解码器60进行优化后的学习模型。如上所述,模型生成单元132可以使用诸如VAE和GAN等各种已知的方法生成模型。

[0112] 获取单元133获取各种信息。例如,获取单元133获取被输入至模型生成单元132生成的学习模型的数据。

[0113] 例如,获取单元133获取从第一编码器50输出的特征量或从第二编码器55输出的特征量。进一步地,获取单元133可以改变所输出的特征量并且获取已获取的信息作为将被输入至学习模型的信息。例如,获取单元133可以通过从已知特征量的先验分布中对已知特征量进行采样而获得新的特征量。

[0114] 进一步地,获取单元133可以从用户使用的信息处理终端中获取任意数据。例如,获取单元133获取构成歌曲的数据。然后,获取单元133可以将所获取的数据输入至学习模型(在这种情况下,分别将相同的数据输入至第一编码器50和第二编码器55)并且获取从每个编码器输出的特征量。

[0115] 内容生成单元134向解码器输入从学习后的第一编码器50或第二编码器55输出的特征量或者通过对基于这些特征量生成的任意特征量进行组合而获得的值,由此生成与第一内容具有相同格式的第二内容作为新内容。应注意,基于从学习后的第一编码器50或第二编码器55输出的特征量而生成的特征量是例如通过从任意编码器输出的特征量的先验

分布进行采样而获得的特征量。

[0116] 例如,内容生成单元134可以基于通过将任意数据输入至学习后的第一编码器50或第二编码器55而输出的特征量生成第二内容。可替代地,如图3中示出的示例,内容生成单元134可以根据通过将所生成的已知特征量(在图3的示例中为特征量 z_1)与通过将任意数据输入至学习后的第一编码器50或第二编码器55而获得的特征量(在图3的示例中为特征量 z_3)进行组合而获得的值(例如,64维矢量),生成第二内容。

[0117] 应注意,如上所述,保持用户所希望的特征的特征量根据用户指定的操作而改变。即,从第一编码器50输出的特征量和从第二编码器55输出的特征量中的哪一特征量保持用户所希望的特征根据用户指定的操作或提取操作的类型而改变。在任意情况下,如果获得与被输入至解码器60的特征量对应的值(例如,如果特征量是64维矢量,则是构成64维矢量的值),内容生成单元134则能够生成与特征量对应的第二内容。

[0118] 此外,内容生成单元134可以将通过对从学习后的第一编码器50输出的特征量与从学习后的第二编码器55输出的特征量进行组合而获得的值输入至解码器60来生成第二内容,而无需使用已知的特征量。例如,内容生成单元134可以基于通过对由获取单元133获得的任意数据(歌曲等)与通过将任意数据输入至学习后的第一编码器50或第二编码器55而输出的每个特征量进行组合而获得的值,来生成第二内容。

[0119] 此外,内容生成单元134可以通过固定组合值之中与任意特征量对应的值并且按顺序改变与其他特征量对应的值,来按顺序生成多个第二内容。例如,当特征量的前半部分表示节奏特征时,内容生成单元134能够通过固定特征量的前半部分并且改变后半部分,来按顺序生成保留其节奏的多首歌曲。

[0120] [1-3.根据实施方式的信息处理步骤]

[0121] 接着,将参考图6和图7对根据实施方式的信息处理步骤进行描述。首先,将参考图6对根据本公开的实施方式的学习处理的流程进行描述。图6是示出根据本公开的实施方式的信息处理步骤的流程图(1)。

[0122] 如图6中示出的,信息处理装置100判断是否已经获取学习数据(内容)(步骤S101)。当未获取学习数据时(步骤S101;否),信息处理装置100等待,直至获取学习数据。

[0123] 另一方面,当获取了学习数据时(步骤S101;是),信息处理装置100提取目标信息(特征信息)(步骤S102)。

[0124] 然后,信息处理装置100将原始学习数据与所提取的信息输入至单独的编码器(步骤S103)。例如,信息处理装置100将原始学习数据输入至第一编码器50并且将所提取的信息输入至第二编码器55。

[0125] 然后,信息处理装置100计算原始学习数据和所提取信息中每一个的特征量(步骤S104)。随后,信息处理装置100将所计算的特征量输入至解码器60(步骤S105)。

[0126] 从而,信息处理装置100使解码器60输出与原始学习数据的格式相同的数据(步骤S106)。然后,信息处理装置100基于输出数据和原始学习数据对模型进行优化(步骤S107)。例如,信息处理装置100将输出数据与原始学习数据进行比较,并且对模型参数进行优化,以减少重构误差。应注意,为了模型优化,信息处理装置100可以重复从步骤S101至步骤S107的处理足够的次数(例如,直至参数稳定在预定的阈值内)。

[0127] 信息处理装置100将学习模型(第一编码器50、第二编码器55以及解码器60)存储

在存储单元120中(步骤S108)。结果,信息处理装置100结束学习处理。

[0128] 接着,将参考图7对根据本公开的实施方式的生成处理的流程进行描述。图7是示出根据本公开的实施方式的信息处理步骤的流程图(2)。

[0129] 如图7中示出的,信息处理装置100获取第一特征量(步骤S201)。例如,第一特征量是学习后的第一编码器50的输出。进一步地,信息处理装置100获取第二特征量(步骤S202)。例如,第二特征量是学习后的第二编码器55的输出。

[0130] 然后,信息处理装置100将通过对所获取的第一特征量和第二特征量进行组合而获得的值输入至解码器60(步骤S203)。

[0131] 随后,信息处理装置100基于从解码器60输出的数据生成第二内容(步骤S204)。

[0132] (2.其他实施方式)

[0133] 除上述各个实施方式之外,可以通过各种不同的形式(变形)执行根据上述各个实施方式的处理。

[0134] [2-1.学习和生成处理的变形]

[0135] 在上述实施方式中,已经描述了包括两个编码器和一个解码器的学习模型。此处,信息处理装置100可以学习具有与上述实施方式的配置不同的配置的各种模型。将参考图8对此进行描述。

[0136] 图8是示出根据本公开的第一变形的学习处理的示例的示图。与图1中示出的示例相比较,图8中示出的示例与图1中示出的示例的不同在于,模型具有第一解码器61和第二解码器62。

[0137] 在图8示出的示例中,如图1中示出的步骤S1,信息处理装置100对数据 x_1 执行提取操作40(步骤S15),并且提取特征信息 f_1 (步骤S16)。进一步地,如图1中示出的步骤S5一样,信息处理装置100将数据 x_1 输入至第一编码器50并且将特征信息 f_1 输入至第二编码器55(步骤S17)。从而信息处理装置100获取特征量 z_1 和特征量 z_2 (步骤S18)。换言之,该操作是指信息处理装置100对数据 x_1 和特征信息 f_1 执行与第一编码器50或第二编码器55对应的功能操作并且输出特征量 z_1 和特征量 z_2 。

[0138] 随后,信息处理装置100将特征量 z_1 和特征量 z_2 输入至第一解码器61并且将特征量 z_2 输入至第二解码器(步骤S19)。换言之,该操作是指信息处理装置100对特征量 z_1 和特征量 z_2 执行与第一解码器61对应的功能操作,并且对特征量 z_2 执行与第二解码器62对应的功能操作。

[0139] 信息处理装置100通过步骤S19中的操作从第一解码器61获得数据 x' 。进一步地,信息处理装置100从第二解码器62获得特征信息 f' (步骤S20)。应注意,数据 x' 是与数据 x_1 具有相同格式的数据并且概念性地表示经由解码器输出的数据。此外,特征信息 f' 是与特征信息 f_1 具有相同格式的数据并且概念性地表示经由解码器输出的特征信息。

[0140] 之后,信息处理装置100对各编码器和解码器的参数进行调整,以使得数据 x' 接近数据 x_1 并且特征信息 f' 接近特征信息 f_1 。即,信息处理装置100通过与图1中示出的步骤S10相同的方式执行模型优化处理。

[0141] 因此,信息处理装置100可以学习具有两个解码器的模型。换言之,信息处理装置100生成具有第一解码器61和第二解码器62的学习模型,第一解码器61基于第一特征量(特征量 z_1)和第二特征量(特征量 z_2)解码出与第一内容(数据 x_1)对应的数据,第二解码器62基

于第二特征量解码出与从构成第一内容的元素中提取的第一数据(特征信息 f_1)对应的数据(特征信息 f_2)。

[0142] 根据该配置,与图1中示出的配置相比较,通过第二编码器55和第二解码器62学习的特征量 z_2 必将具有关于特征信息 f_1 的信息。出于此原因,信息处理装置100能够可靠地执行特征分离的学习。更具体地,信息处理装置100通过图8中示出的配置,能够更为可靠地学习被分离为特征信息 f_1 的信息(例如,歌曲的节奏)。

[0143] 应注意,图8中示出的模型能够具有又一不同的配置。将参考图9对此进行描述。图9是示出根据本公开的第二变形的学习处理的示例的示意图。

[0144] 与图8中示出的示例相比较,图9中示出的示例与图8中示出的示例的不同在于,模型不具有第二编码器55。

[0145] 在图9示出的示例中,如图8中示出的步骤S15,信息处理装置100对数据 x_1 执行提取操作40(步骤S31),并且提取特征信息 f_1 (步骤S32)。此处,信息处理装置100将数据 x_1 输入至第一编码器50(步骤S33)并且获得特征量 z_1 和特征量 z_2 (步骤S34)。在这种情况下,信息处理装置100针对从第一编码器50输出的值提前设置预定规则并且输出特征量 z_1 和特征量 z_2 。例如,信息处理装置100提前设置规则,以使得所输出的64维矢量的前32维是特征量 z_1 并且后32维是特征量 z_2 。

[0146] 随后,信息处理装置100将特征量 z_1 和特征量 z_2 输入至第一解码器61,并且将特征量 z_2 输入至第二解码器62(步骤S35)。信息处理装置100通过步骤S35中的操作从第一解码器61获得数据 x' 。进一步地,信息处理装置100从第二解码器62获得特征信息 f' (步骤S36)。

[0147] 之后,信息处理装置100对每个编码器和解码器的参数进行调整,以使得数据 x' 接近数据 x_1 并且特征信息 f' 接近特征信息 f_1 。

[0148] 如上所述,即使在其中省去第二编码器55的配置中,信息处理装置100也能够执行如图1和图8示出的对分离特征量的学习。因此,信息处理装置100能够简化模型的配置,结果,例如能够快速执行学习处理或能够减少处理负荷。

[0149] 接着,将参考图10描述生成新内容的情况下的生成处理的变形。图10是示出根据本公开的第三变形的生成处理的示例的示意图。

[0150] 图10示出了在从学习后的第一编码器50和第一解码器61生成新的内容时的处理的流程。应注意,例如,第一编码器50和第一解码器61是通过图9中的学习等其参数经过学习(训练)的编码器和解码器。

[0151] 信息处理装置100将任意数据 x_1 输入至第一编码器50(步骤S41)并且获得特征量 z_1 和特征量 z_2 (步骤S42)。然后,信息处理装置100将特征量 z_1 和特征量 z_2 输入至第一解码器61(步骤S43)并且获得数据 x' (步骤S44)。应注意,信息处理装置100可以不从第一编码器50获得被输入至第一解码器61的特征量,而是通过从事先学习的特征量的分布中对被输入至第一解码器61的特征量进行采样来获得特征量。

[0152] 以这种方式,信息处理装置100能够从具有简单配置的学习模型(在本示例中,为第一编码器50和第一解码器61)生成作为新内容的数据 x' 。根据该配置,信息处理装置100将特征量 z_1 和特征量 z_2 输入至一个解码器并且因此能够获得例如包括未被完全分离的特征的数据 x' 。结果,信息处理装置100能够生成甚至包括节奏信息的新内容,该节奏信息未被完全分离成例如特征量 z_2 并且还被假设甚至保留在特征量 z_1 中。

[0153] [2-2.提取操作的变形]

[0154] 在上述实施方式中,作为提取操作40的示例,已经描述了其中执行预定操作从歌曲30中提取顺序不变信息(例如,节奏信息)的示例。除本示例之外,信息处理装置100还可以执行各种操作作为提取操作40。

[0155] 例如,在上述实施方式中,已经示出了其中信息处理装置100提取歌曲30的和弦构成声音的示例,但是,信息处理装置100不仅可以提取和弦构成声音,而且还可以提取旋律的构成声音或鼓的构成声音。例如,旋律的构成声音表示诸如歌曲的小节等某个部分中出现的声音类型等。此外,例如,鼓的构成声音表示诸如歌曲的小节等某个部分中出现的鼓组等的声音类型(小鼓、低音鼓等)。

[0156] 此外,信息处理装置100可以提取表示音高的流的信息(所谓的旋律轮廓(Melodic Contour)),诸如在歌曲的流中音高是上升、下降、还是停留等。此外,信息处理装置100可以提取构成歌曲的声音的音阶(表示12种声音中的哪种声音被用于创作歌曲的信息)。进一步地,信息处理装置100可以提取赋予构成声音的顺序的信息(有顺序的构成声音)。例如,信息处理装置100可以提取其中相邻音符的音高进行排序和布置的信息(例如,诸如1度、1度以及2度等音高的序列)。

[0157] 此外,信息处理装置100可以提取歌曲中的模式分类序列(例如,隐含/实现模型等)。进一步地,信息处理装置100可以采用提取通过例如调性音乐的生成理论(GTMM)的时间跨度树的深度所确定的重要音符序列的方法。

[0158] 此外,信息处理装置100可以从构成歌曲的声音中提取主音(lead sound)、和弦音、和弦类型、和弦的基音、低音线等。

[0159] 进一步地,当歌曲由MIDI信息构成时,信息处理装置100可以提取某个部分中的强度信息(velocity)。

[0160] 进一步地,信息处理装置100可以提取构成歌曲的每个乐器或每个乐器组的声音。此外,当通过深层神经网络(DNN)学习某首歌曲的特征时,信息处理设备100可以提取通过计算某层的特征之间的相关性而获得的风格特征量等。进一步地,信息处理装置100可以提取歌曲中的自相似性。

[0161] 此外,信息处理装置100可以学习歌曲的特征并且基于学习结果提取特征。例如,信息处理装置100可以学习通过序列到序列网络(模型)对歌曲进行转调的任务并且对解码器设定诸如转调等条件,以提取对于转调不变的特征。

[0162] [2-3.应用变形]

[0163] 如上述实施方式中所述,根据本公开的信息处理,可以执行其中分离预定特征的学习,因此,能够选择性地增加或减少新内容中所出现的特征的影响。将参考图11对此进行描述。图11是用于对根据本公开的生成处理进行概念性地描述的示图。

[0164] 图11示出了概念性地示出通过根据本公开的生成处理生成的歌曲的图像(音高和音长)的图表64。如上所述,信息处理装置100能够提取节奏作为表示歌曲的一个特征并且生成反映固定或改变节奏的影响的新歌曲。

[0165] 例如,在固定确定歌曲的节奏特征的因子(被称为“节奏因子”)的同时改变除节奏因子之外的特征时,信息处理装置100沿着图11中示出的固定节奏因子的轴而生成歌曲。另一方面,当固定除节奏因子之外的特征并且改变节奏因子时,信息处理装置100沿着图11中

示出的固定除节奏因子之外的轴而生成歌曲。

[0166] 通常,在内容生成中,当不能分离特征量时,则可以对原始数据进行一些改变,但是难以调整改变的方向。例如,信息处理装置100能够通过改变歌曲65的表示1小节的语音信息的特征量而生成新歌曲66,但是,当不能分离特征量时,改变的方向将被设置成对角线上的一个方向(所谓的1D插值)。

[0167] 然而,根据信息处理装置100,因为能够从歌曲的特征中分离出诸如节奏因子等元素,所以能够在其中固定节奏因子的方向上和其中固定除节奏因子之外的元素的方向(所谓的2D插值)上改变歌曲65。例如,信息处理装置100能够根据用户需求将歌曲65改变成图表64中示出的所有歌曲的图像。如上所述,信息处理装置100能够生成新的内容,以对特征量的混合比进行调整。

[0168] 即,信息处理装置100不仅能够通过固定歌曲的节奏或转调程度、音阶等,而且还能够通过控制改变的程度来生成新内容或变形。作为一种特定的方法,能够通过获得两种噪声并且将每种噪声添加至两个特征量 z_1 和 z_2 可生成特征量的变化。此时,当噪声被缩放时,能够控制两个特征量 z_1 和 z_2 中的每个特征量的变化程度。例如,当存在用于获得噪声的两种方法时,存在(1)从诸如正常分布等某个固定分布中获得噪声的方法和(2)使用VAE来学习编码器并且使用从编码器输出的噪声的方法。此外,信息处理装置100能够通过交换某两首歌曲的特征而执行诸如生成新内容等灵活的生成处理。

[0169] [2-4.特征量的变形]

[0170] 在上述实施方式中,已经描述了其中信息处理装置100学习歌曲中的两种类型的特征量(节奏及除节奏之外的元素)的示例。然而,信息处理装置100可以学习两种或更多种类型的特征量。

[0171] 将参考图12对此进行描述。图12是示出根据本公开的第四变形的学习处理的示例的示图(1)。图12概念性地示出了在学习处理中分离的特征量 z_o 、 z_a 、 z_b 和 z_c 以及特征量之间的关系变化。

[0172] 图12的(a)中示出的概念图67示出了三种类型的特征量是排他特征的特征组。图12的(b)中示出的概念图68示出了四种类型的特征量是作为共同部分的某个空间的特征组。图12的(c)中示出的概念图69示出了三种类型的特征量是作为嵌套空间的特征组。

[0173] 图13示出了学习图12的(a)中示出的特征组的示例。图13是示出根据本公开的第四变形的学习处理的示例的示图(2)。

[0174] 如图13中示出的,信息处理装置100对数据 x_1 执行第一提取操作41和第二提取操作42(步骤S51)。然后,信息处理装置100通过第一提取操作41获得特征信息 f_1 并且通过第二提取操作42获得特征信息 f_2 (步骤S52)。随后,信息处理装置100将数据 x_1 输入至第一编码器50,将特征信息 f_1 输入至第二编码器55,并且将特征信息 f_2 输入至第三编码器56(步骤S53)。

[0175] 信息处理装置100分别从第一编码器50获得特征量 z_o ,从第二编码器55获得特征量 z_a ,并且从第三编码器56获得特征量 z_b (步骤S54)。随后,信息处理装置100分别将特征量 z_o 、特征量 z_a 以及特征量 z_b 输入至第一解码器61,将特征量 z_a 输入至第二解码器62并且将特征量 z_b 输入至第三解码器63(步骤S55)。然后,信息处理装置100从第一解码器61获得数据 x' ,从第二解码器62获得特征信息 f_1' ,并且从第三解码器63获得特征信息 f_2' (步骤S56)。随

后,信息处理装置100对解码器和编码器进行优化,分别使得数据 x' 接近数据 x_1 ,特征信息 f_1' 接近特征信息 f_1 并且特征信息 f_2' 接近特征信息 f_2 。

[0176] 通过图13中示出的配置,信息处理装置100能够学习具有图12的(a)中示出的关系的特征组。应注意,作为图13中示出的学习的特定示例,例如,第一解码器61执行与歌曲的旋律有关的任务,第二解码器62执行与歌曲的旋律的节奏有关的任务,并且第三解码器63执行与歌曲的旋律构成声音有关的任务。换言之,在图13示出的模型中,可以说第二解码器62捕获特征量 z_a ,第三解码器63捕获特征量 z_b ,并且第一解码器61捕获其他特征。

[0177] 接着,图14示出了学习图12的(b)中示出的特征组的示例。图14是示出根据本公开的第四变形的学习处理的示例的示意图(3)。

[0178] 如图14中示出的,信息处理装置100将数据 x_1 输入至第一编码器50(步骤S61)。然后,信息处理装置100从第一编码器50输出特征量 z_o 、 z_a 、 z_b 以及 z_c (步骤S62)。应注意,在这种情况下,信息处理装置100可以通过与图9中的步骤S33相同的方式设置关于特征量的分离的预定规则等。

[0179] 随后,信息处理装置100分别将特征量 z_o 、 z_a 、 z_b 以及 z_c 输入至第一解码器61,将特征量 z_a 和特征量 z_c 输入至第二解码器62,并且将特征量 z_b 和 z_c 输入至第三解码器63(步骤S63)。如图12的(b)中示出的,该处理基于事实:特征量 z_c 与特征量 z_a 以及与特征量 z_b 具有共同的特征。

[0180] 然后,信息处理装置100从第一解码器61获得数据 x' ,从第二解码器62获得特征信息 f_1' ,并且从第三解码器63获得特征信息 f_2' (步骤S56)。随后,信息处理装置100对解码器和编码器进行优化,以分别使得数据 x' 接近数据 x_1 ,特征信息 f_1' 接近特征信息 f_1 ,并且特征信息 f_2' 接近特征信息 f_2 。

[0181] 通过图14中示出的配置,信息处理装置100能够学习具有图12的(b)中示出的关系的特征组。

[0182] 接着,图15示出了学习图12的(c)中示出的特征组的示例。图15是示出根据本公开的第四变形的学习处理的示例的示意图(4)。

[0183] 在图15示出的处理中,步骤S71至S74与图13中示出的步骤S51至S54是相同的,并且由此省去其描述。

[0184] 信息处理装置100分别将在步骤S74获得的特征量 z_o 、特征量 z_a 以及特征量 z_b 输入至第一解码器61,将特征量 z_a 和特征量 z_b 输入至第二解码器62,并且将特征量 z_b 输入至第三解码器63(步骤S75)。如图12的(c)中示出的,该处理基于事实:特征量 z_b 与特征量 z_a 和特征量 z_o 具有嵌套结构。

[0185] 然后,信息处理装置100分别从第一解码器61获得数据 x' ,从第二解码器62获得特征信息 f_1' ,并且从第三解码器63获得特征信息 f_2' (步骤S76)。随后,信息处理装置100对解码器和编码器进行优化,以分别使得数据 x' 接近数据 x_1 ,特征信息 f_1' 接近特征信息 f_1 ,并且特征信息 f_2' 接近特征信息 f_2 。

[0186] 通过图15中示出的配置,信息处理装置100能够学习具有图12的(c)中示出的关系的特征组。如参考图12至图15描述的,即使当多个特征量(特征组)具有不同的关系时,信息处理装置100也能够通过分离每个特征而执行学习。

[0187] [2-5.信息处理的变形]

[0188] 当两个特征量(例如,特征量 z_1 和特征量 z_2)之间存在依赖关系时,信息处理装置100可以依赖于每次处理所使用的编码器或先验分布。

[0189] 进一步地,信息处理装置100可以在编码器的两个通道中共享嵌入矢量(embedding)。例如,当信息处理装置100通过提取歌曲的构成声音等而执行学习时,信息处理装置100在编码器的两个通道中共享嵌入矢量。

[0190] 进一步地,信息处理装置100可以使用敌对学习分离特征量。具体地,信息处理装置100可以学习对特征量 z_2 和正常分布的噪声进行分类的标识符,以分离两个特征量(例如,特征量 z_1 和特征量 z_2),并且可以执行诸如学习特征量 z_1 等敌对学习,使得识别变得不能进行。

[0191] [2-6.模型结构的变形]

[0192] 信息处理装置100可以学习关于特征量或特征信息的处理具有另外不同结构的模型。下面将参考图16对此进行描述。图16是示出根据本公开的第五变形的学习处理的示例的示意图。

[0193] 在图16示出的示例中,信息处理装置100对数据 x_1 执行提取操作40(步骤S81)并且提取特征信息 f_1 (步骤S82)。之后,信息处理装置100将特征信息 f_1 与数据 x_1 一起输入至第一编码器50并且将特征信息 f_1 输入至第二编码器55(步骤S83)。因此,信息处理装置100获得特征量 z_1 和特征量 z_2 (步骤S84)。

[0194] 随后,信息处理装置100将特征量 z_1 和特征信息 f_1 输入至第一解码器61,并且将特征量 z_2 输入至第二解码器62(步骤S85)。然后,信息处理装置100从第一解码器61获得数据 x' 并且从第二解码器62获得特征信息 f' (步骤S86)。

[0195] 之后,信息处理装置100对每个编码器和解码器的参数进行调整,以使得数据 x' 接近数据 x_1 并且特征信息 f' 接近特征信息 f_1 。即,信息处理装置100执行模型优化处理。

[0196] 如上所述,信息处理装置100可以生成具有第一解码器61和第二解码器62的学习模型,第一解码器61基于第一特征量(特征量 z_1)和第一数据(特征信息 f_1)来解码出与第一内容(数据 x_1)对应的数据,第二解码器62基于第二特征量(特征量 z_2)解码出与第一数据对应的数据(特征信息 f')。

[0197] 具体地,不同于图8中示出的模型等,在步骤S85,信息处理装置100将特征信息 f_1 输入至第一解码器61。因此,在对特征信息 f_1 进行调节之后,信息处理装置100能够执行编码器和解码器的学习。即,信息处理装置100能够执行包括特征信息 f_1 的学习,由此排除与特征信息 f_1 有关的不确定性。具体地,信息处理装置100能够提前执行包括特征信息 f_1 的学习,因此,可以防止由于某些因子而出现不执行分离特征量的学习的情形(例如,仅提取歌曲的节奏信息作为特征量等)。

[0198] 更具体地,根据图16中的模型的结构,不针对特征量 z_1 学习特征信息 f_1 的不确定性,并且仅学习数据 x_1 的除特征信息 f_1 之外的部分的不确定性。换言之,对于特征量 z_1 ,学习与特征信息 f_1 无关的空间。另一方面,通过其他编码器(在图16的示例中,为第二编码器55)学习与特征信息 f_1 有关的空间。结果,学习特征量 z_1 和特征量 z_2 作为明确分离的空间。即,根据第五变形的学习处理,信息处理装置100能够提高学习稳定性。

[0199] 接着,将参考图17描述生成新内容的情况下的生成处理的变形。图17是示出根据本公开的第五变形的生成处理的示例的示意图(1)。

[0200] 图17示出了从学习后的第一编码器50、第一解码器61以及第二解码器62生成新内容时的处理的流程。应注意,例如,第一编码器50、第一解码器61以及第二解码器62是通过图16中的学习而学习其参数的编码器和解码器。

[0201] 信息处理装置100对任意数据 x_1 执行提取操作40(步骤S87)并且提取特征信息 f_1 (步骤S88)。之后,信息处理装置100将特征信息 f_1 与数据 x_1 一起输入至第一编码器50(步骤S89)。作为结果,信息处理装置100获得特征量 z_1 (步骤S90)。此外,信息处理装置100通过从事前学习的特征量(空间)的分布中采样事前学习的特征量(空间)而获得任意特征量 z_2 。

[0202] 随后,信息处理装置100将所获得的特征量 z_2 输入至第二解码器62(步骤S91)并且将所获得的特征信息 f' 、特征量 z_1 以及特征信息 f_1 输入至第一解码器61(步骤S91)。因此,信息处理装置100获得数据 x' ,即,新内容(步骤S92)。

[0203] 通过图17中的配置,信息处理装置100能够生成反映通过提取操作提取的特征(在图17的示例中为特征信息 f_1)的影响的数据 x' 作为新内容。

[0204] 应注意,信息处理装置100在生成时可以采用又一不同的配置。将参考图18对此进行描述。

[0205] 图18示出了从学习后的第一编码器50、第二编码器55、第一解码器61以及第二解码器62生成新内容时的处理的流程。应注意,例如,第一编码器50、第二编码器55、第一解码器61以及第二解码器62是通过图16中的学习而学习其参数的编码器和解码器。

[0206] 信息处理装置100对任意数据 x_1 执行提取操作40(步骤S93)并且提取特征信息 f_1 (步骤S94)。之后,信息处理装置100将特征信息 f_1 与数据 x_1 一起输入至第一编码器50,并且将特征信息 f_1 输入至第二编码器55(步骤S95)。因此,信息处理装置100从第一编码器50获得特征量 z_1 并且从第二编码器55获得特征量 z_2 (步骤S96)。此时,信息处理装置100可通过从事前学习的特征量(空间)的分布中采样事前学习的特征量(空间)而获得任意特征量 z_1 。

[0207] 随后,信息处理装置100将所获得的特征量 z_2 输入至第二解码器62(步骤S97)并且将所获得的特征信息 f' 和特征量 z_1 输入至第一解码器61。因此,信息处理装置100获得数据 x' ,即,新内容(步骤S98)。

[0208] 通过图18中的配置,信息处理装置100能够生成反映通过提取操作提取的特征(在图18的示例中,为特征信息 f_1)的影响的数据 x' 作为新内容。

[0209] [2-7.内容的变形]

[0210] 在上述实施方式中,内容的示例包括歌曲(音乐)。此处,根据变形的信息处理装置100可以使用图像、文本数据等作为内容而执行根据本公开的信息处理。

[0211] 将参考图19至图21描述其中信息处理装置100将图像作为内容进行处理的示例。图19是示出根据本公开的第六变形的信息处理的示例的示图(1)。

[0212] 例如,当将图像作为内容进行处理时,信息处理装置100执行从图像中提取高分辨率信息或低分辨率信息的操作。

[0213] 例如,假设信息处理装置100学习图19中示出的作为内容的图像80并且生成学习模型。假设构成图像80的像素的数量是“ 128×128 像素”。在这种情况下,在诸如“ 32×32 像素”等的占据一定大小的粗略范围内聚集(pooled,合并出)每个像素信息。结果,获得了对于每个“ 32×32 像素”,图像80由一个像素表示的低分辨率图像(通过对图像80进行马赛克处理而获得的图像)。

[0214] 然后,信息处理装置100利用第一编码器50学习图像80的原始数据并且利用第二编码器55学习低分辨率图像。然后,从第一编码器50输出的特征量 z_1 是表示比 32×32 像素更小的范围内的特征的特征量。另一方面,从第二编码器55输出的特征量 z_2 是表示除由特征量 z_1 表示的特征之外的更宽范围内的特征的特征量。

[0215] 例如,由图像80中的 32×32 像素表示的特征不与人对应,而是与诸如背景等宽范围的颜色对应。另一方面,例如,表示图像80中比 32×32 像素更小的范围内的特征对应于例如与诸如人的眼睛或鼻子的部位等更小范围内的特征的信息等。

[0216] 出于此原因,信息处理装置100在固定特征量 z_1 的同时改变特征量 z_2 (例如,在固定特征量 z_1 的同时,利用先验分布对特征量 z_2 进行采样),并且当生成多个第二内容时,能够生成图19中示出的图像组82。图19中示出的图像组82中所包括的每个图像是其中诸如图像80中包括的人的眼睛或鼻子等细微部位几乎不改变并且仅背景颜色和环境光发生改变的图像。

[0217] 按照这种方式生成的图像组82是指通过在各种背景和环境对同一人成像而获得的多个图像。即,根据信息处理装置100,可以从对某个人成像而获得的一个图像80中获得与在不同环境中对同一人成像而获得的那些图像相当的大量图像。因此,信息处理装置100能够从非常小数量的原始图像获得用于生成执行特定处理的机器学习模型(例如,面部识别或面部表情识别等分类器)的较大数量的图像。换言之,信息处理装置100能够通过适当的方式使得学习所使用的图像增加。

[0218] 进一步地,信息处理装置100能够执行与图19中的学习不同的学习。将参考图20对不同学习的示例进行描述。图20是示出根据本公开的第六变形的信息处理的示例的示图(2)。

[0219] 作为学习的原始数据,使用与图19相同的图像80。在图20的示例中,在占有与图19相比更小范围的范围(诸如图像80的“ 8×8 像素”等)内聚集每个像素信息。

[0220] 然后,信息处理装置100利用第一编码器50学习图像80的原始数据并且利用第二编码器55学习聚集图像。然后,从第一编码器50输出的特征量 z_1 是表示比 8×8 像素更小的范围内的特征的特征量。另一方面,从第二编码器55输出的特征量 z_2 是表示除由特征量 z_1 表示的特征之外的更宽范围内的特征的特征量。

[0221] 例如,图像80中的比 8×8 像素更小的范围内所表示的特征与例如诸如人的眼睛和鼻子的部位等更小范围内的特征的信息等对应。

[0222] 在图20的示例中,信息处理装置100在固定特征量 z_2 的同时改变特征量 z_1 (例如,在固定特征量 z_2 的同时,利用先验分布对特征量 z_1 进行采样),并且当生成多个第二内容时,能够生成图20中示出的图像组87。不同于图19,图20中示出的图像组87中所包括的每个图像是其中背景颜色或环境光几乎不发生改变并且诸如图像80中包括的人的眼睛或鼻子等细微部位发生改变的图像。

[0223] 例如,按照这种方式生成的图像组87能够是用于生成高度准确的面部识别模型的较大数量的图像。即,当在图像80中捕获的人发生细微地改变时,可以使用图像组87生成做出不对在图像80中错误地捕获的人进行识别的判断的模型。因此,信息处理装置100能够改善面部识别模型的性能。

[0224] 应注意,信息处理装置100能够使用又一不同的学习方法。将参考图21对此进行描

述。图21是示出根据本公开的第六变形的信息处理的示例的示图(3)。

[0225] 作为学习的原始数据,使用与图19相同的图像80。此处,在图21的示例中,在对图像80进行编码时,使用在中间层出现的数据来使第二编码器55学习。例如,信息处理装置100对第二编码器55的中间层的数据(例如,4×4像素的数据)执行全局均聚集(GAP)操作,并使第二编码器55学习,以输出特征量 z_2 。因此,在特征量 z_2 中,学习位置信息(背景等)已经消失的图像80,并且在特征量 z_1 中,学习位置信息。在这种情况下,通过固定特征量 z_1 并且对特征量 z_2 进行采样可以操作除图像中的位置信息之外的信息。例如,如图21中示出的,生成图像组92,其中,保留背景的大致氛围,但是图像中包括的人的面部或人的发型在各个方面是不同的。换言之,信息处理装置100能够利用面部的改变特征属性(对准)生成包括各种人的各种图像。根据该处理,信息处理装置100能够从例如一个图像中生成贴有“人”的标签的较大数量的学习图像。

[0226] 如上所述,信息处理装置100可以对构成第一内容的数字数据(即,图像)执行提取操作并且提取第一数据。例如,信息处理装置100通过对图像的特定区域中所包括的每个像素的像素信息执行具有顺序不变性的运算处理(例如,聚集)作为提取操作而提取第一数据。

[0227] 如上所述,即使内容不是歌曲而是图像,信息处理装置100也能够执行其中分离用户作为目标的特征的学习。此外,信息处理装置100能够使用所生成的学习模型自动生成保留用户所指定的特征的图像,诸如其中固定某个特征(诸如背景或人的面部等)并且改变其他特征的图像等。

[0228] 此外,信息处理装置100可以使用文本数据、动画等作为内容。即,如果可以通过单独编码器学习经由某些操作提取的信息和原始内容数据,信息处理装置100则能够执行分离目标特征量的学习,而与内容的类型无关。

[0229] 进一步地,信息处理装置100可以应用根据本公开的信息处理,以基于人类味觉信息执行检测发出类似味道的物质等的处理。例如,信息处理装置100从某种物质中分离出具有某种味道的元素,并且将通过每种元素进行数字化而获得的信息输入至各个编码器,由此学习通过从某种物质中仅分离出特定味道的特征而获得的特征。

[0230] 此外,在上述实施方式中,已经描述了其中信息处理装置100通过执行具有顺序不变性的某种操作或者丢失某种信息来执行提取特征信息的操作的示例。然而,提取操作并不局限于本示例。例如,信息处理装置100基于针对歌曲的每个小节而提前人工设置的“歌曲的兴奋程度”,学习诸如哪一阶段具有较高的“兴奋程度等”的监督回归函数。然后,信息处理装置100基于学习结果计算待处理的歌曲的每个小节的兴奋程度。然后,信息处理装置100仅提取其兴奋程度超过预定阈值的小节并且利用第二编码器55学习所提取的小节。因此,信息处理装置100能够生成可以分离诸如“兴奋程度”的人工设置的特征的学习模型。

[0231] 即,信息处理装置100不一定必须通过算法运算从内容中提取特征信息,而是还能够基于人工设置特征从内容中提取特征信息。利用该配置,例如,因为能够分离与情绪对应的特征,所以可以自动生成例如具有人类所感觉的某些特征的内容,诸如产生高兴情绪的内容等。作为示例,信息处理装置100对自然语言处理应用该配置,以自动生成具有引起某种情绪的特征的内容(句子),诸如“产生幸福情绪的句子”等。

[0232] [2-8.其他]

[0233] 进一步地,在上述实施方式描述的每个处理之中,被描述为自动执行的全部或部分处理可手动执行或者被描述为手动执行的处理可通过已知方法自动执行。此外,除非另有规定,否则,能够任意改变上述文献和附图中示出的处理步骤、具体名称以及包括各种数据和参数的信息。例如,每个附图中示出的各种信息并不局限于所示出的信息。

[0234] 此外,所示出的每个设备的每个部件是功能性概念并且并不一定需要如图所示在物理上配置。即,每个设备的分配/集成的具体形式并不局限于附图中示出的形式,并且根据各种负荷和使用条件,可以将其全部或部分功能地或物理地分配/集成到任意单元中。

[0235] 进一步地,能够将上述实施方式和变形适当地组合在不与处理内容冲突的范围内。

[0236] 此外,本说明书中描述的效果仅是示出性的并且并不局限于例证并且可以存在额外的效果。

[0237] (3. 硬件配置)

[0238] 例如,通过具有图11中示出的配置的计算机1000实现诸如根据上述各个实施方式的信息处理装置100等信息设备。在下文中,将根据实施方式的信息处理装置100描述为示例。图22是示出实现信息处理装置100的功能的计算机1000的示例的硬件配置图。计算机1000包括CPU 1100、RAM 1200、只读存储器(ROM) 1300、硬盘驱动器(HDD) 1400、通信接口1500以及输入/输出接口1600。计算机1000的每个部分通过总线1050连接。

[0239] CPU 1100基于存储在ROM 1300或HDD 1400中的程序而操作并且控制各个单元。例如,CPU 1100将存储在ROM 1300或HDD 1400中的程序扩展到RAM 1200中并且执行与各种程序对应的处理。

[0240] ROM 1300存储诸如当计算机1000启动时通过CPU 1100运行的基本输入输出系统(BIOS)等启动程序、取决于计算机1000的硬件的程序等。

[0241] HDD 1400是非临时地记录通过CPU 1100运行的程序、程序所使用的数据等的计算机可读记录介质。具体地,HDD 1400是记录根据本公开的信息处理程序(作为程序数据1450的示例)的记录介质。

[0242] 通信接口1500是用于将计算机1000连接至外部网络(例如,互联网)的接口。例如,CPU 1100经由通信接口1500从其他设备接收数据并且将通过CPU 1100生成的数据发送至其他设备。

[0243] 输入/输出接口1600是用于在输入/输出设备1650与计算机1000之间连接的接口。例如,CPU 1100经由输入/输出接口1600从诸如键盘或鼠标等输入设备接收数据。此外,CPU 1100经由输入/输出接口1600将数据发送至诸如显示器、扬声器、或打印机等输出设备。进一步地,输入/输出接口1600可以用作读取预定记录介质上所记录的程序等的介质接口。例如,介质是诸如数字通用盘(DVD)或相变重写盘(PD)等光学记录介质、诸如磁光盘(MO)、磁带介质、磁记录介质等磁光记录介质、半导体存储器等。

[0244] 例如,当计算机1000用作根据实施方式的信息处理装置100时,计算机1000的CPU 1100通过执行被加载到RAM 1200上的信息处理程序而实现控制单元130的功能等。进一步地,HDD 1400将根据本公开的信息处理程序或数据存储在存储单元120中。应注意,CPU 1100从HDD 1400读取并且运行程序数据1450,但作为另一示例,CPU 1100可以经由外部网络1550从其他设备获取这些程序。

- [0245] 应注意,本技术还能够被配置成如下。
- [0246] (1) 一种信息处理装置,包括:
- [0247] 提取单元,从构成第一内容的元素中提取第一数据;和
- [0248] 模型生成单元,生成具有第一编码器和第二编码器的学习模型,第一编码器计算作为第一内容的特征量的第一特征量,第二编码器计算作为所提取的第一数据的特征量的第二特征量。
- [0249] (2) 根据(1)所述的信息处理装置,其中,
- [0250] 模型生成单元,
- [0251] 生成具有解码器的学习模型,解码器基于第一特征量和第二特征量对与第一内容对应的数据进行解码。
- [0252] (3) 根据(1)或(2)所述的信息处理装置,其中,
- [0253] 提取单元,
- [0254] 接收用户针对第一内容而指定的操作和对执行操作的范围的指定并且在所接收的范围内执行操作。
- [0255] (4) 根据(3)所述的信息处理装置,其中,
- [0256] 提取单元,
- [0257] 通过对范围内包括的元素执行具有顺序不变性的运算处理作为操作来提取第一数据。
- [0258] (5) 根据(3)或(4)所述的信息处理装置,其中,
- [0259] 提取单元,
- [0260] 通过丢弃范围内包括的元素中与时间无关或与顺序无关的信息而提取第一数据。
- [0261] (6) 根据(3)~(5)中任一项所述的信息处理装置,其中,
- [0262] 提取单元,
- [0263] 对构成歌曲的第一内容的数字数据执行操作并且提取第一数据。
- [0264] (7) 根据(6)所述的信息处理装置,其中,
- [0265] 提取单元,
- [0266] 执行丢失第一内容中与范围内包括的声音信息中的声音的高度有关的信息的操作并且提取第一数据。
- [0267] (8) 根据(3)~(5)中任一项所述的信息处理装置,其中,
- [0268] 提取单元,
- [0269] 对构成图像的第一内容的数字数据执行操作并且提取第一数据。
- [0270] (9) 根据(8)所述的信息处理装置,其中,
- [0271] 提取单元,
- [0272] 通过对图像的特定区域中所包括的每个像素的像素信息执行具有顺序不变性的运算处理作为该操作而提取第一数据。
- [0273] (10) 根据(2)~(9)中任一项所述的信息处理装置,进一步包括:
- [0274] 内容生成单元,通过将从学习后的第一编码器或第二编码器输出的特征量或组合基于特征量生成的特征量中任一特征量而获得的值输入至解码器来生成第二内容,第二内容是与第一内容具有相同格式的新内容。

- [0275] (11) 根据(10)所述的信息处理装置,其中,
- [0276] 内容生成单元,
- [0277] 通过将从学习后的第一编码器输出的特征量与从学习后的第二编码器输出的特征量进行组合而获得的值输入至解码器而生成第二内容。
- [0278] (12) 根据(10)或(11)所述的信息处理装置,其中,
- [0279] 内容生成单元,
- [0280] 通过按顺序固定与组合的值之中的任意一个特征量对应的值并且改变与其他特征量对应的值,来按顺序生成多个第二内容。
- [0281] (13) 一种信息处理方法,包括:
- [0282] 允许计算机,
- [0283] 从构成第一内容的元素中提取第一数据;并且
- [0284] 生成具有第一编码器和第二编码器的学习模型,第一编码器计算作为第一内容的特征量的第一特征量,并且第二编码器计算作为所提取的第一数据的特征量的第二特征量。
- [0285] (14) 一种用于运行计算机用作下列项的程序,
- [0286] 提取单元,从构成第一内容的元素中提取第一数据;和
- [0287] 模型生成单元,生成具有第一编码器和第二编码器的学习模型,第一编码器计算作为第一内容的特征量的第一特征量,并且第二编码器计算作为所提取的第一数据的特征量的第二特征量。
- [0288] (15) 根据(1)或(3)~(12)中任一项所述的信息处理装置,其中,
- [0289] 模型生成单元,
- [0290] 生成具有第一解码器和第二解码器的学习模型,第一解码器基于第一特征量和第二特征量对与第一内容对应的数据进行解码,第二解码器基于第二特征量对与第一数据对应的数据进行解码。
- [0291] (16) 根据(1)或(3)~(12)中任一项所述的信息处理装置,其中,
- [0292] 模型生成单元,
- [0293] 生成具有第一解码器和第二解码器的学习模型,第一解码器基于第一特征量和第一数据对与第一内容对应的数据进行解码,第二解码器基于第二特征量对与第一数据对应的数据进行解码。
- [0294] 符号说明
- [0295] 100 信息处理装置
- [0296] 110 通信单元
- [0297] 120 存储单元
- [0298] 121 模型存储单元
- [0299] 122 歌曲数据存储单元
- [0300] 130 控制单元
- [0301] 131 提取单元
- [0302] 132 模型生成单元
- [0303] 133 获取单元

[0304] 134 内容生成单元

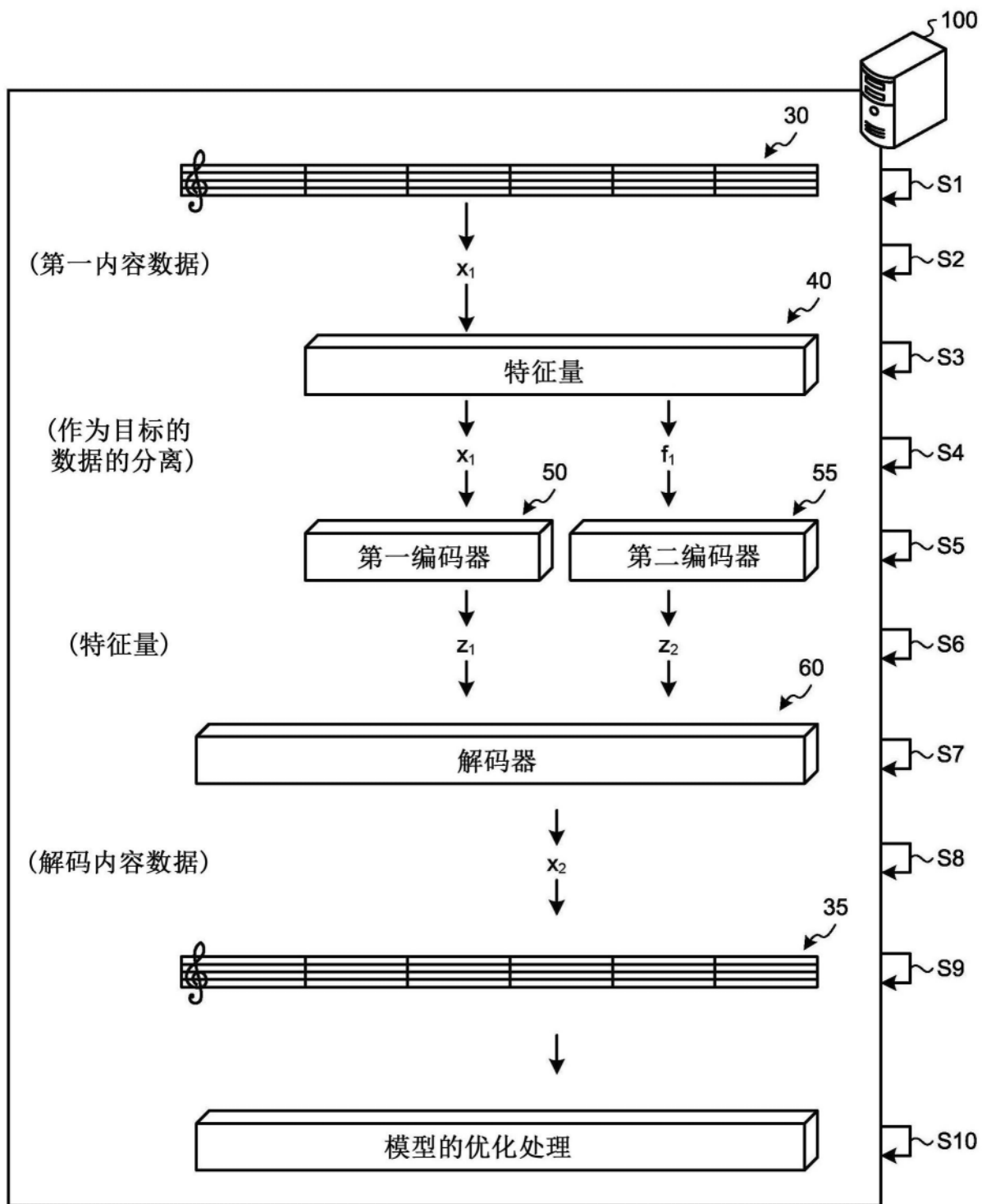


图1

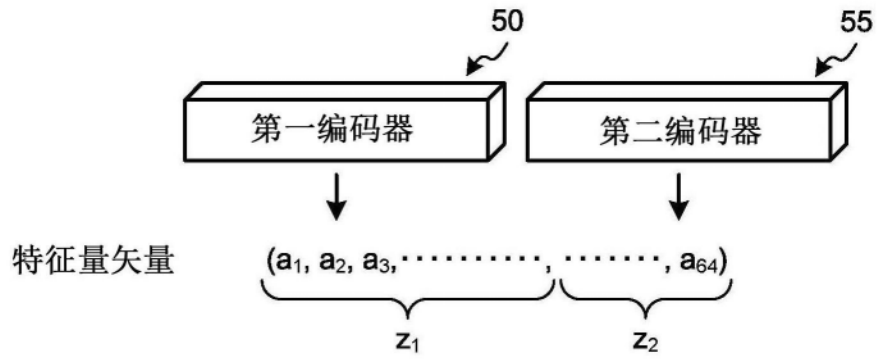


图2

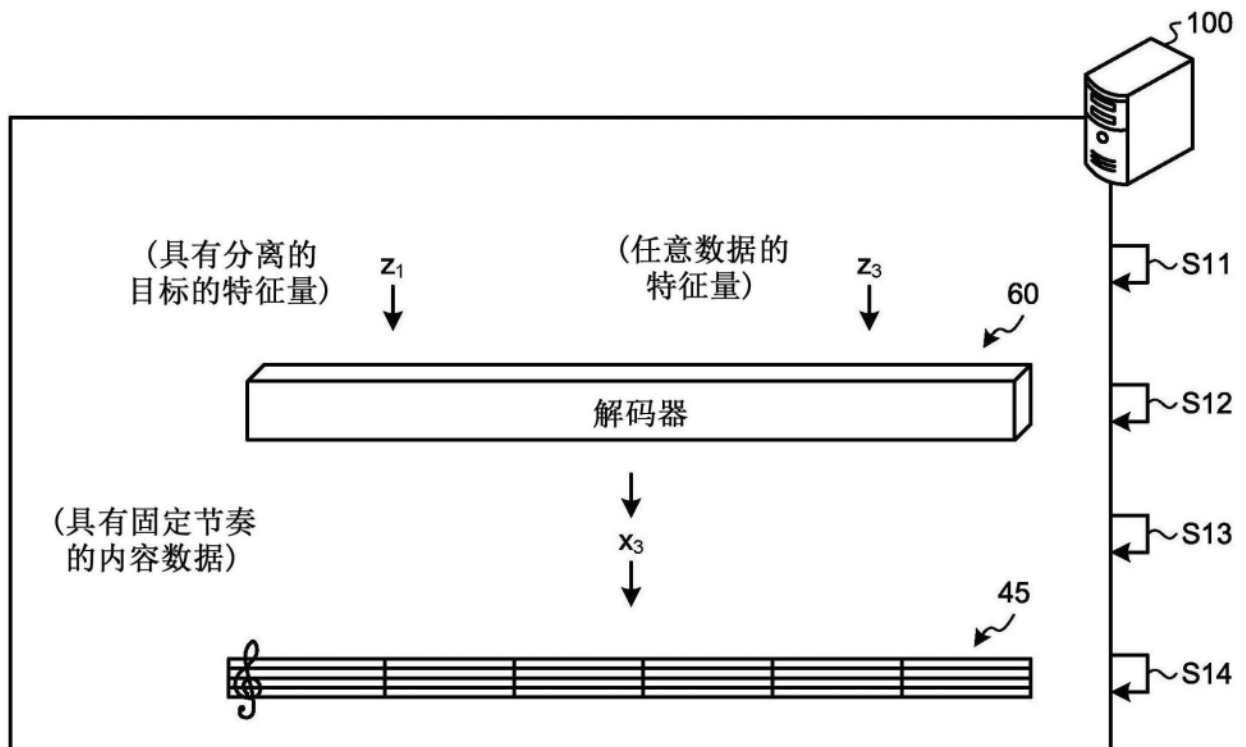


图3

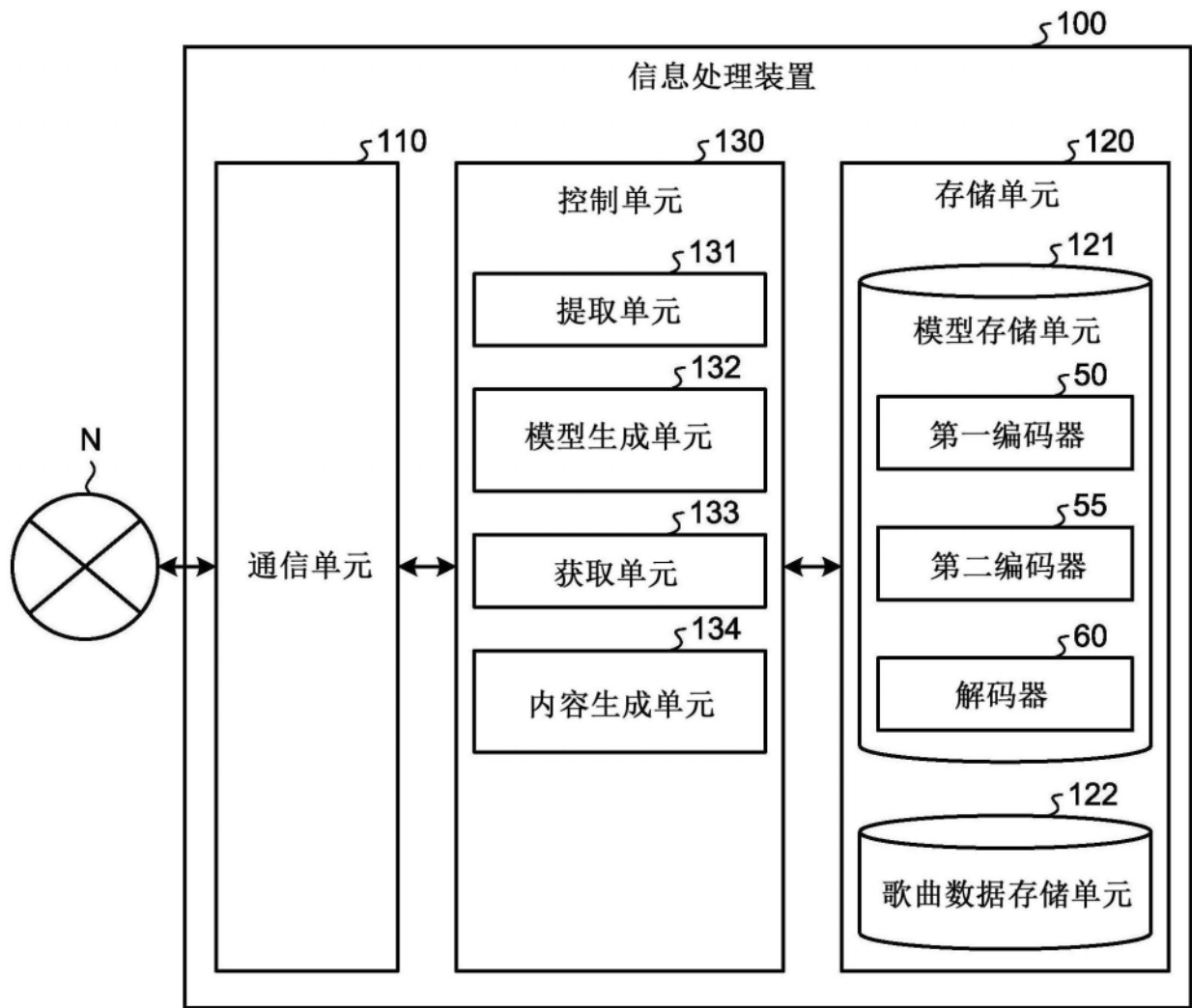


图4

122
↙

歌曲ID	部分数据ID	音高信息	音长、休止符信息	和弦信息	节奏信息	...
...	...	...	...	...	...	...
A01	B01	C01	D01	E01	F01	...
	B02	C02	D02	E02	F02	...
	...	...	...	...	...	...
...	...	...	...	...	...	...

图5

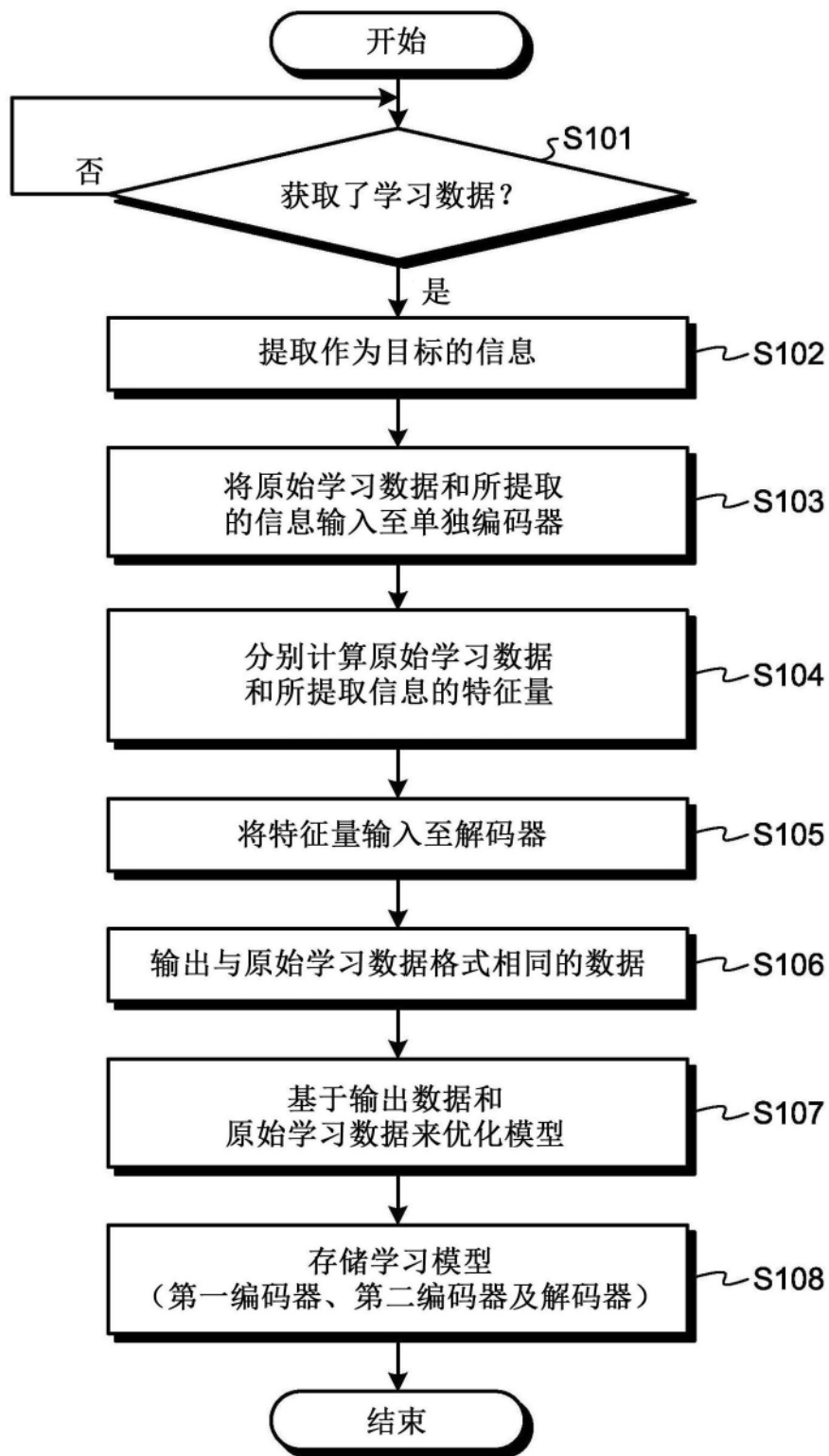


图6

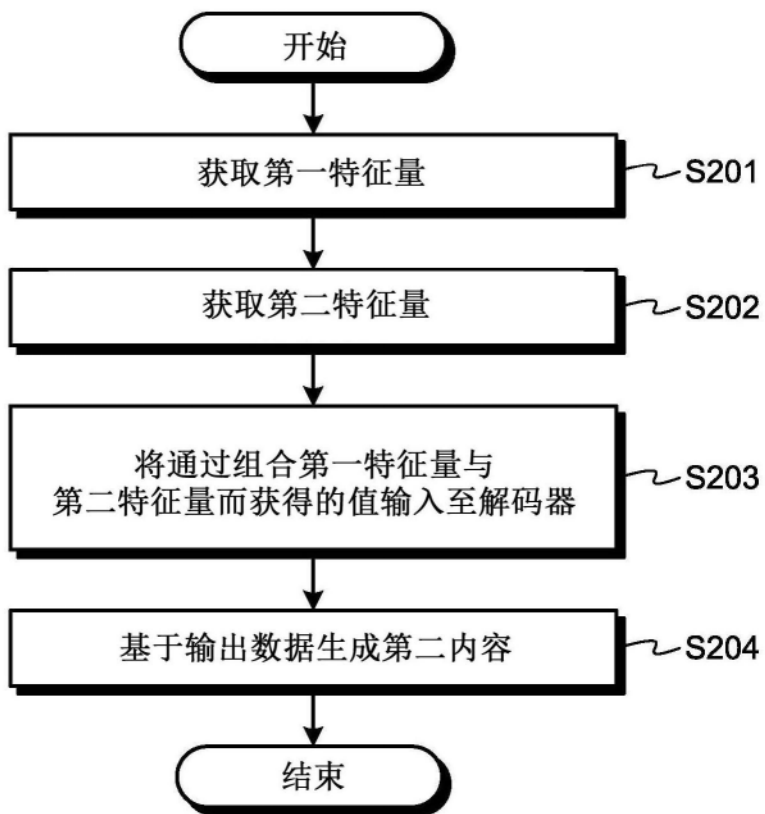


图7

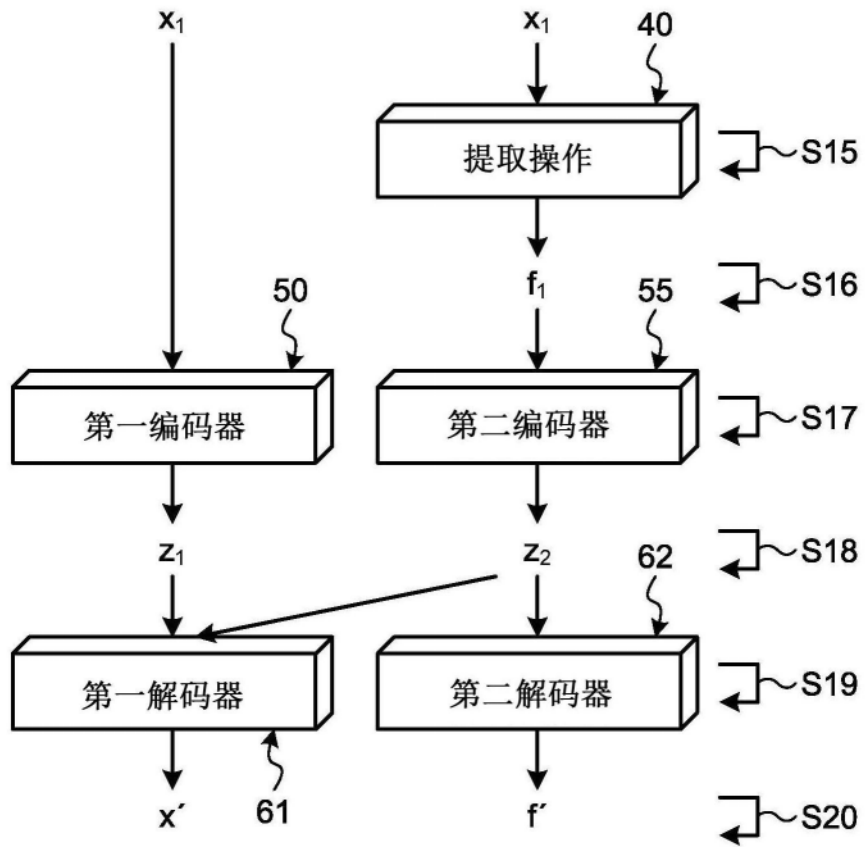


图8

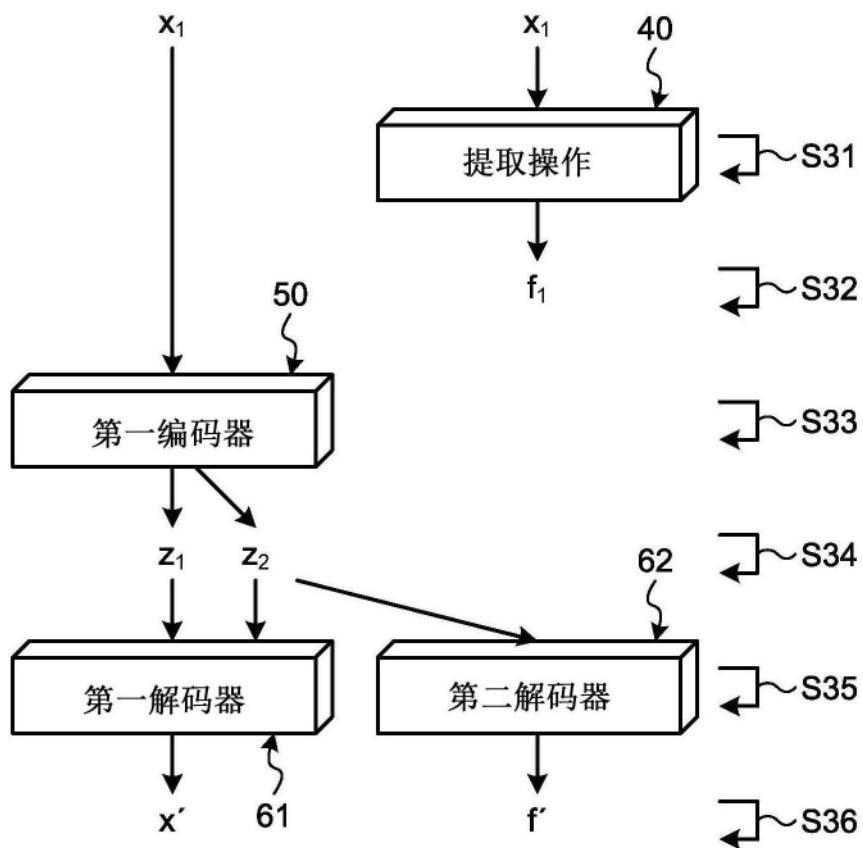


图9

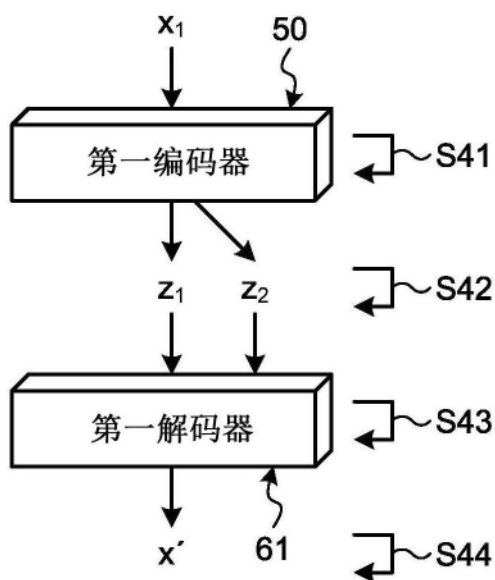


图10

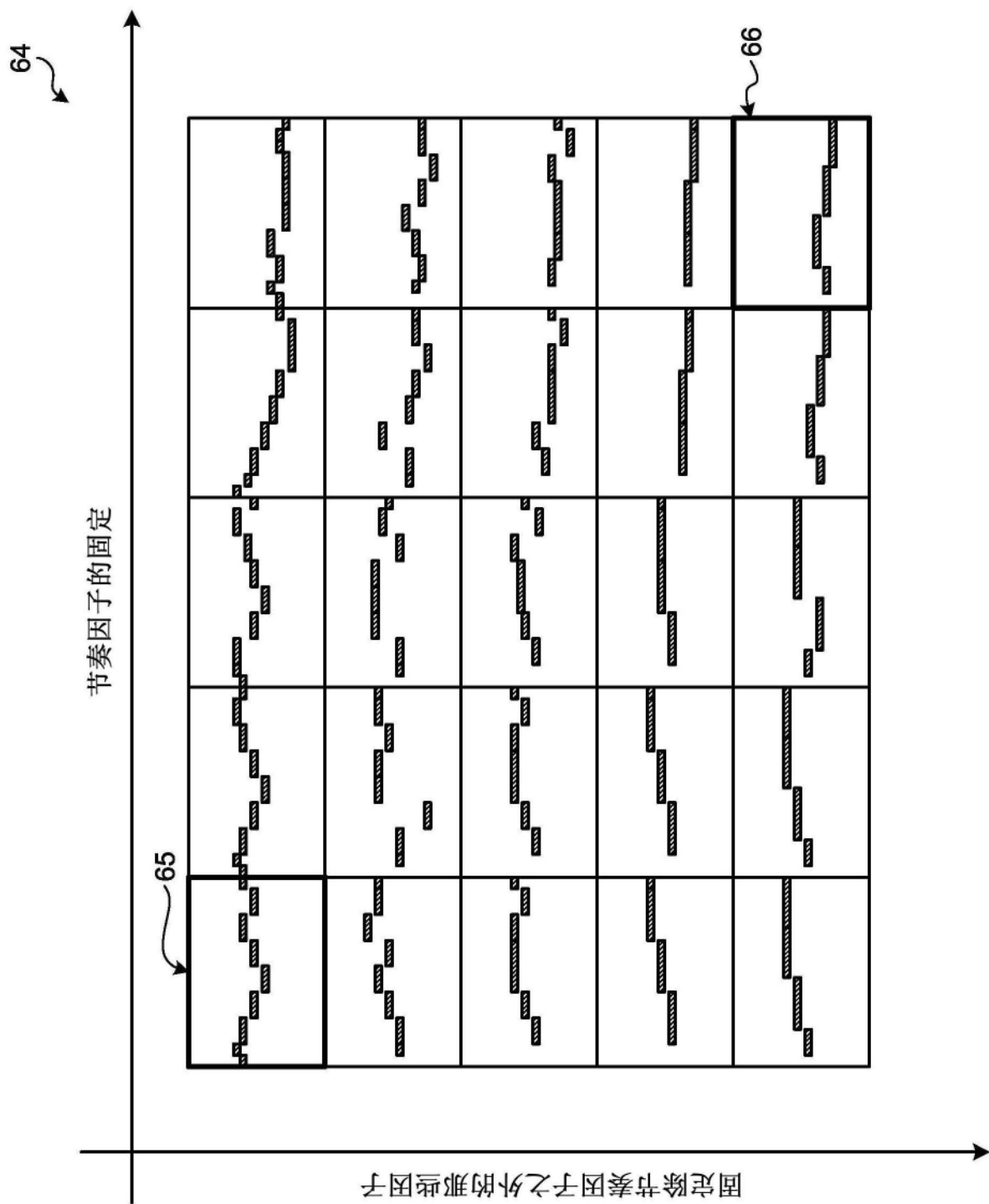
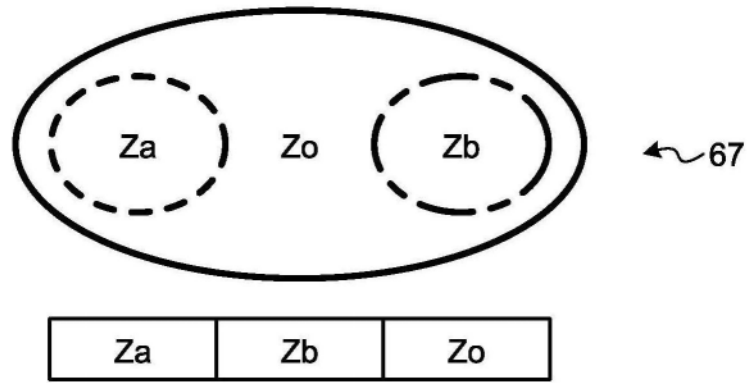
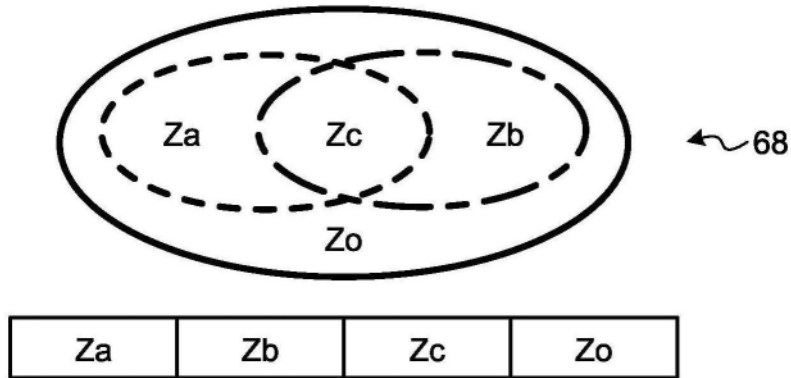


图11

(a)



(b)



(c)

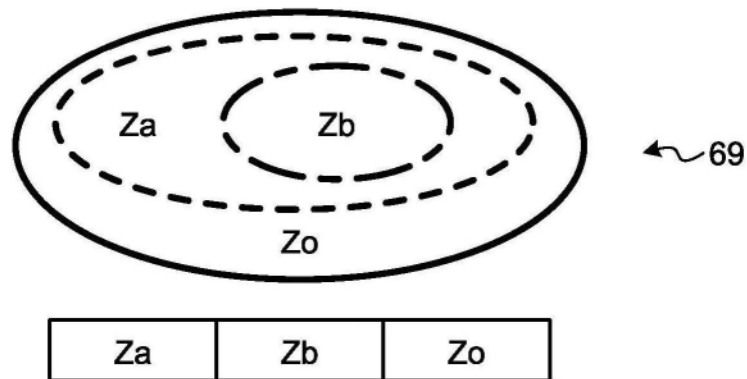


图12

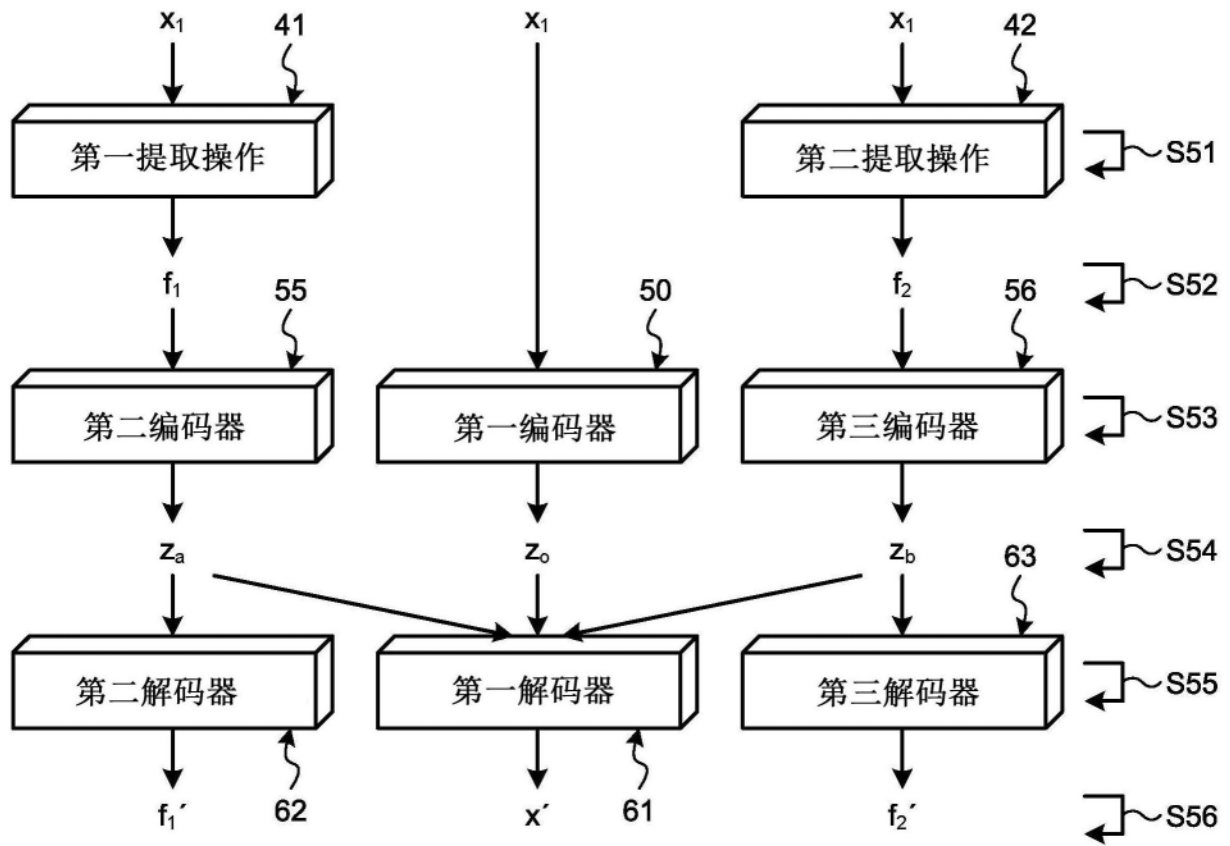


图13

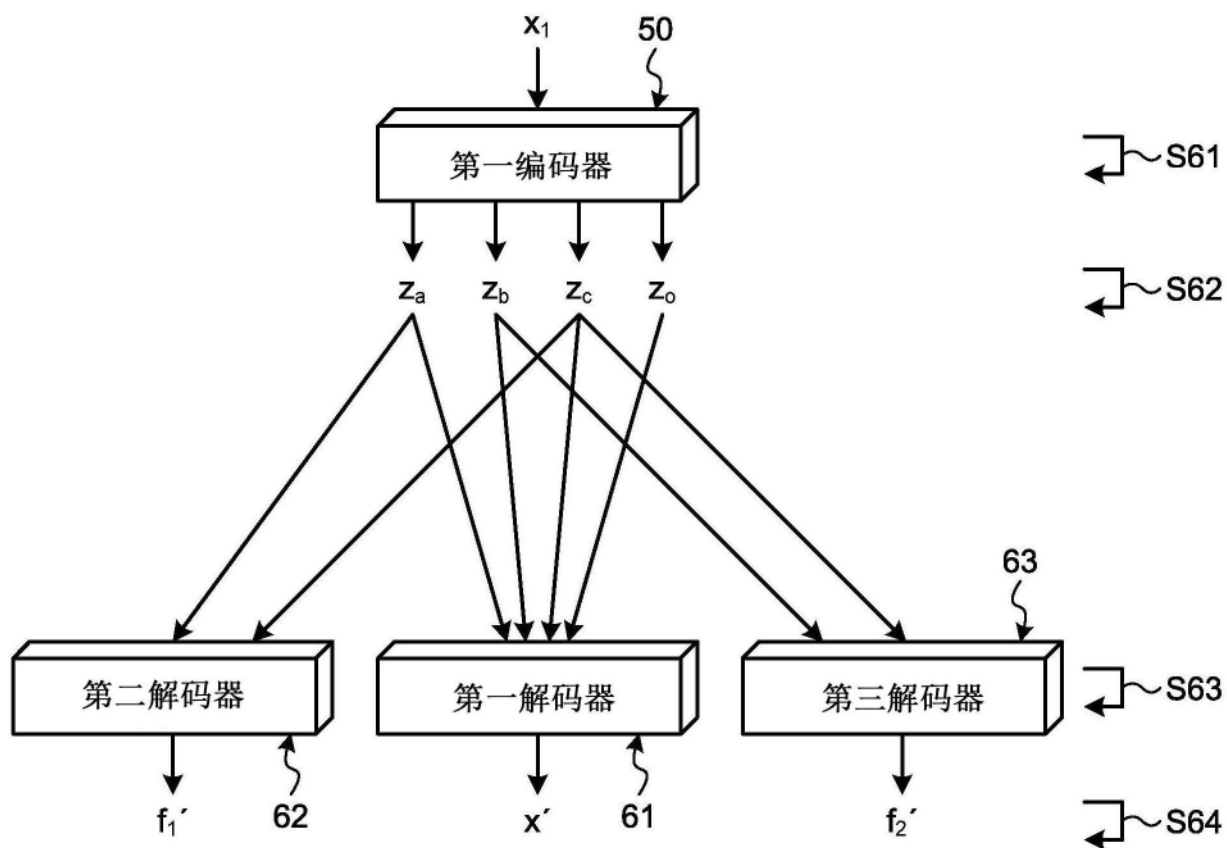


图14

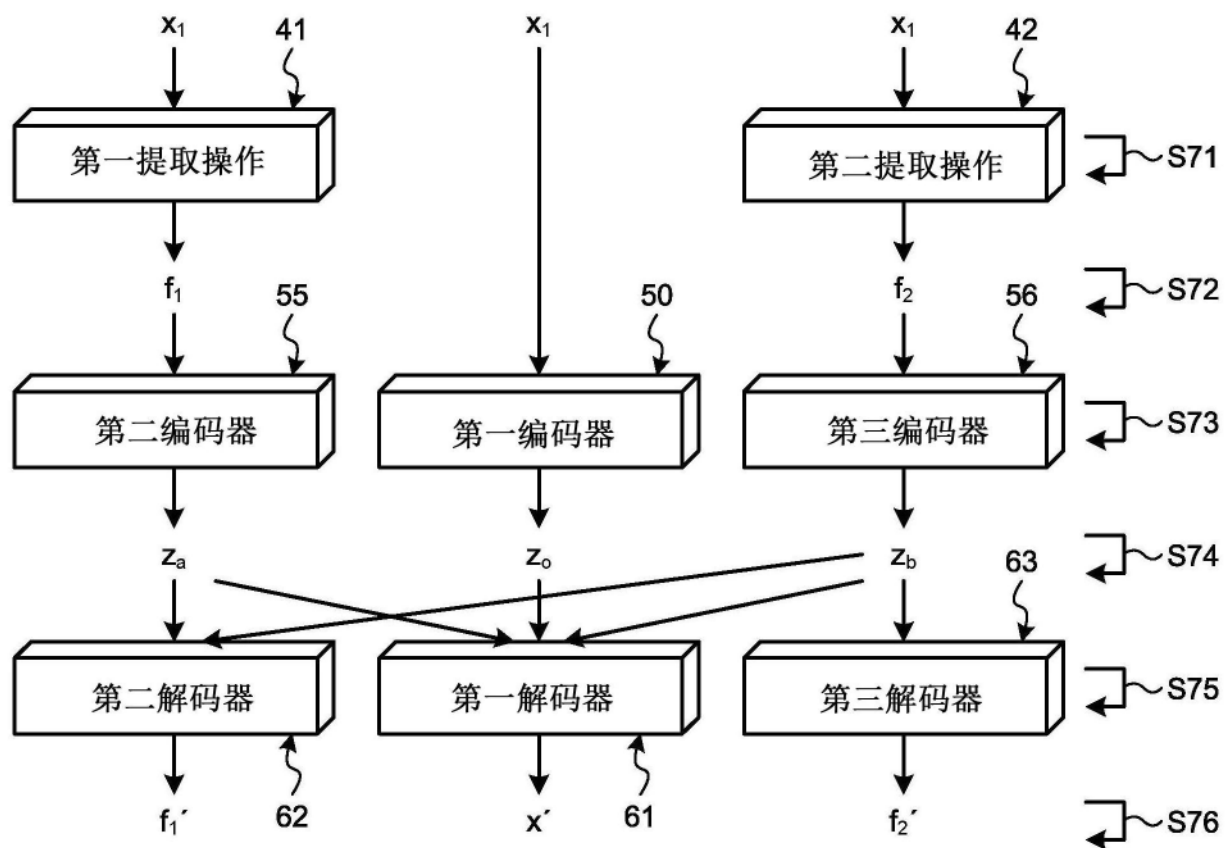


图15

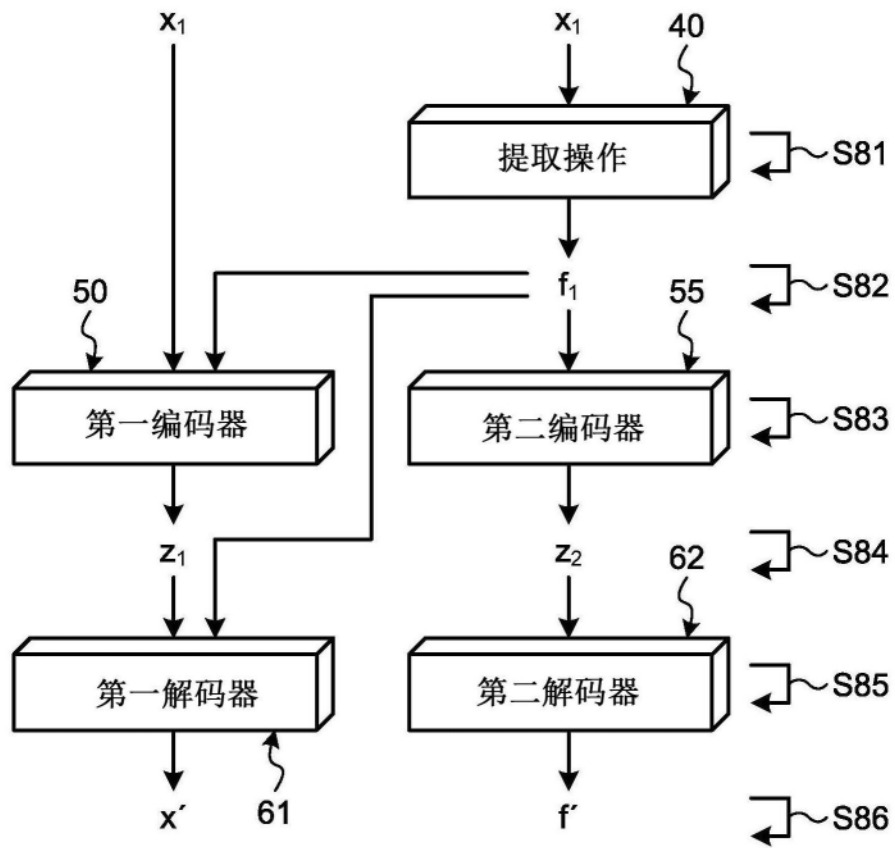


图16

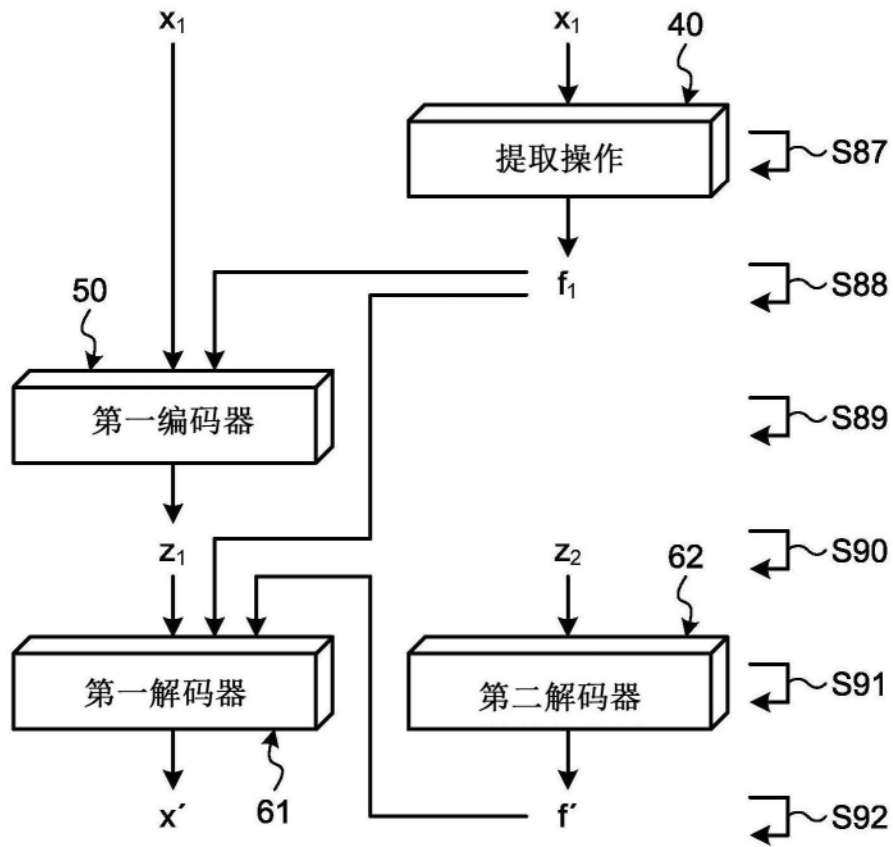


图17

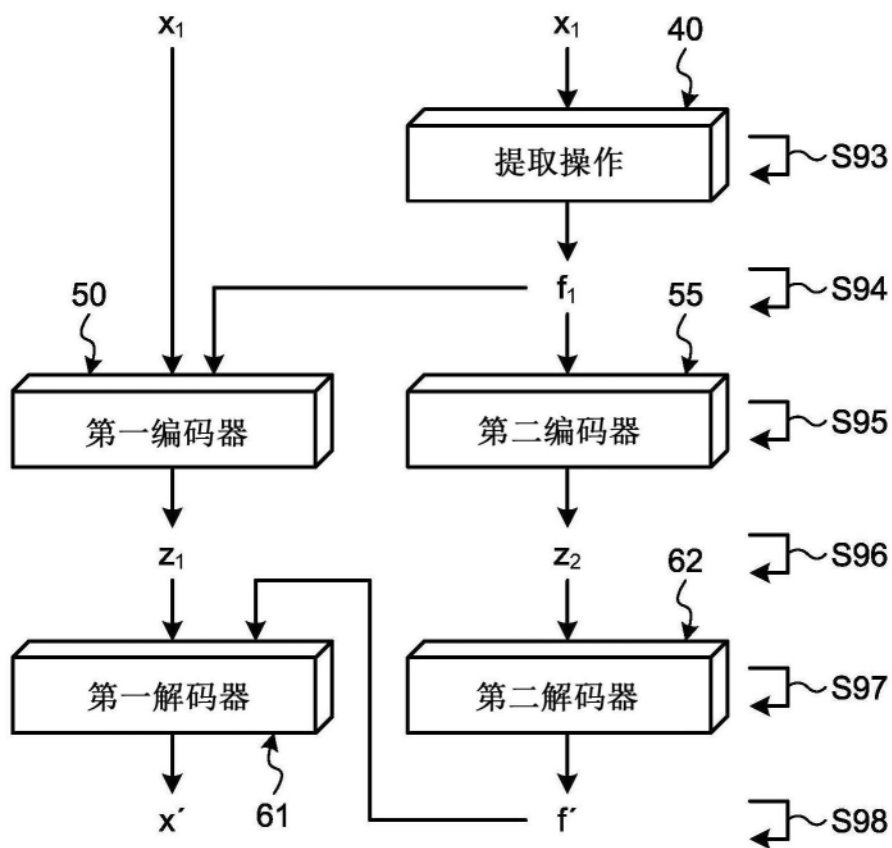


图18

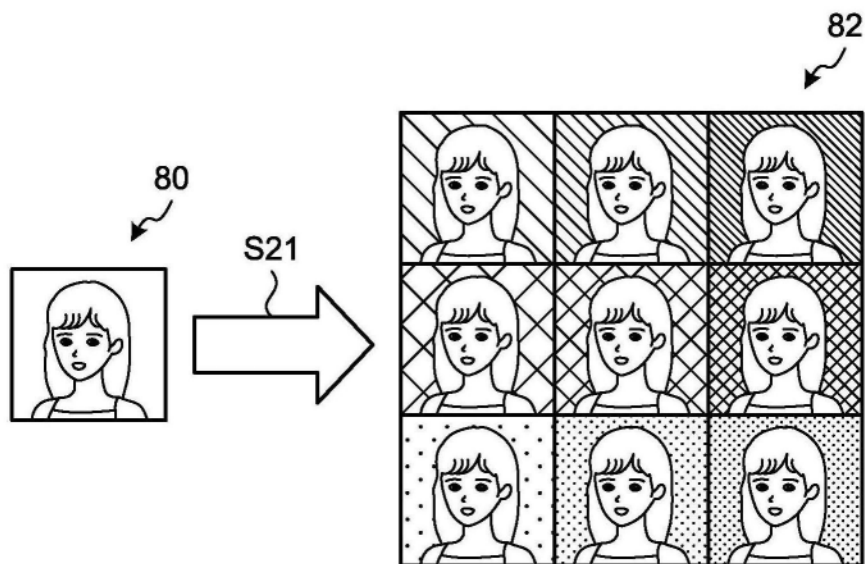


图19

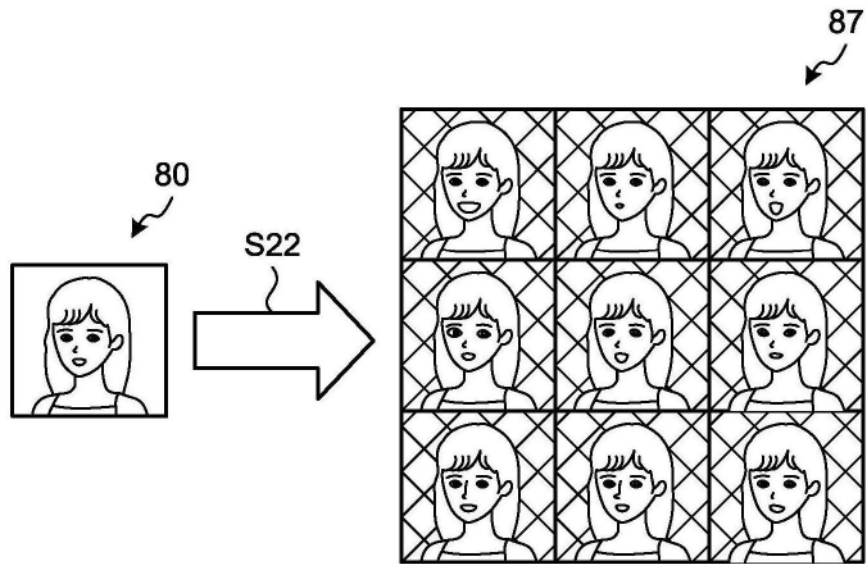


图20

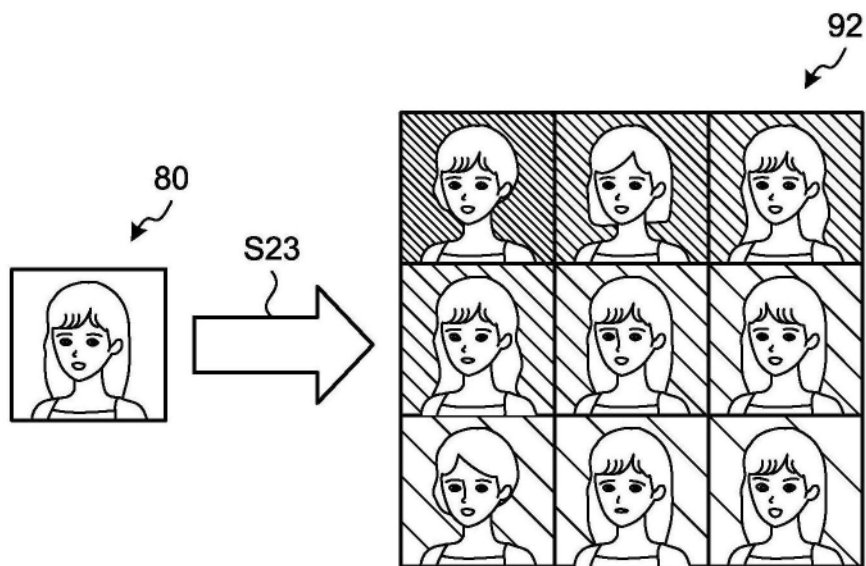


图21

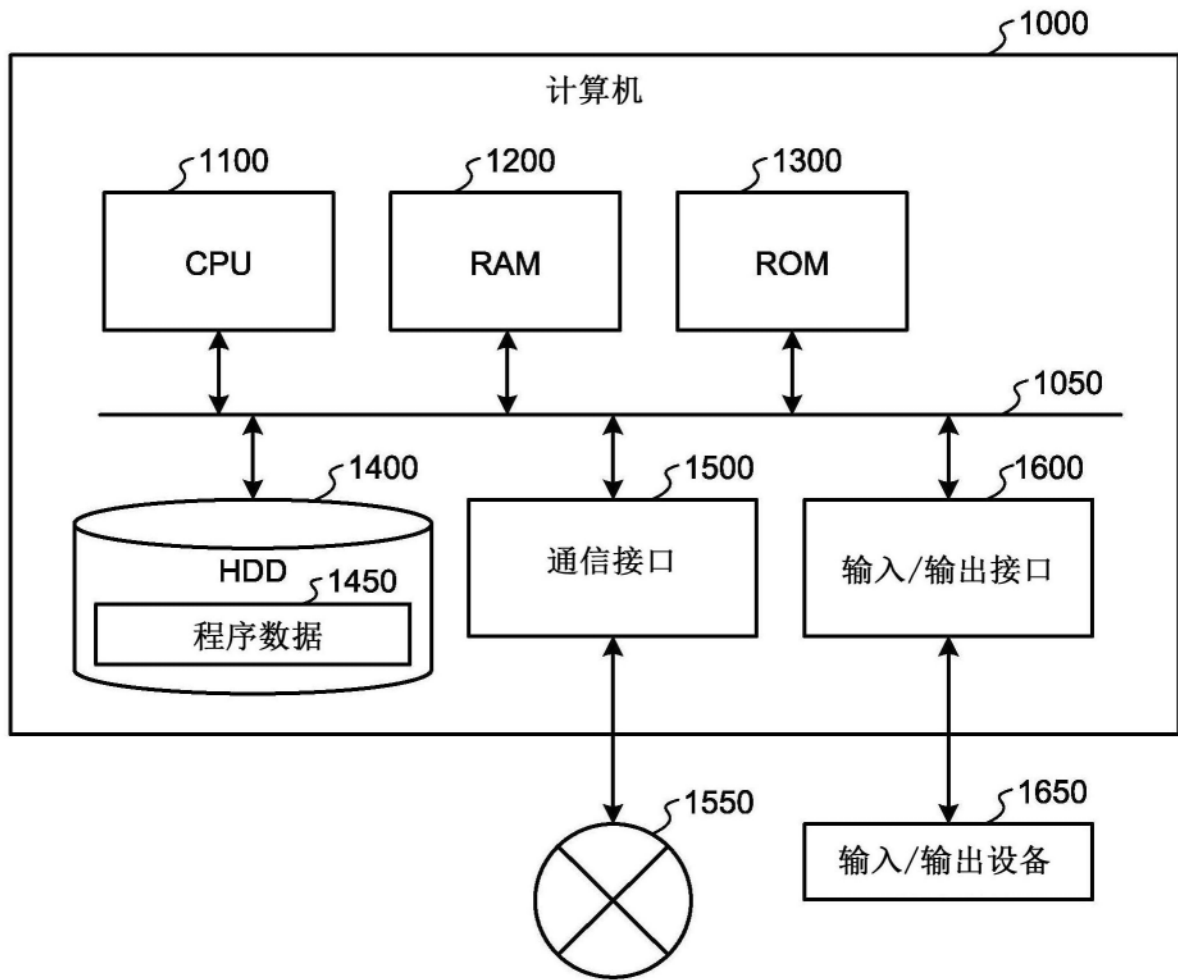


图22