(19)国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 116018613 A
(43)申请公布日 2023.04.25

(21)申请号 202080105040.6
(22)申请日 2020.09.25
(85)PCT国际申请进入国家阶段日
2023.03.10
(86)PCT国际申请的申请数据
PCT/JP2020/036456 2020.09.25
(87)PCT国际申请的公布数据
WO2022/064660 JA 2022.03.31
(71)申请人 富士通株式会社
地址 日本神奈川县
(72)发明人 斋藤淳哉 内田昭嘉 村濑健太郎
(74)专利代理机构 北京集佳知识产权代理有限公司 11227
专利代理师 王海奇
(51)Int.Cl.
G06T 7/00(2017.01)

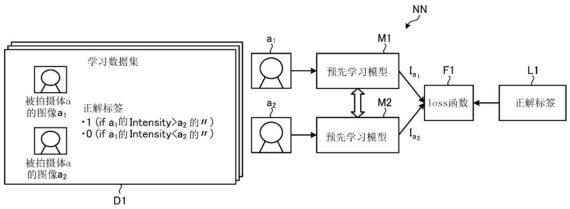
权利要求书2页 说明书10页 附图9页

(54)发明名称

机器学习程序、机器学习方法以及推定装置

(57)摘要

实施方式的机器学习程序使计算机执行生成训练完毕模型的处理和生成第三模型的处理。在生成训练完毕模型的处理中，根据包含第一图像和第二图像的组以及第一标签的训练数据的输入，执行第一模型的机器学习，从而生成训练完毕模型，第一标签表示第一图像和上述第二图像中的哪个图像是被拍摄体的表情肌的动作大的图像，第一模型的机器学习基于将第一图像输入到第一模型而得到的第一输出值、将第二图像输入到与第一模型共享参数的第二模型而得到的第二输出值以及第一标签。在生成第三模型的处理中，通过基于将第三图像输入到训练完毕模型而得到的第三输出值以及第二标签的机器学习，从而生成第三模型，第二标签表示第三图像所包含的被拍摄体的表情肌的动作的产生强度或产生的有无。



CN 116018613 A

1.一种机器学习程序,其特征在于,

使计算机执行如下处理:

根据包含第一图像和第二图像的组以及第一标签的训练数据的输入,执行第一模型的机器学习,从而生成训练完毕模型,其中,所述第一标签表示所述第一图像和所述第二图像中的哪个图像是被拍摄体的表情肌的动作大的图像,所述第一模型的机器学习基于将所述第一图像输入到第一模型而得到的第一输出值、将所述第二图像输入到与所述第一模型共享参数的第二模型而得到的第二输出值、以及所述第一标签,

通过基于将第三图像输入到所述训练完毕模型而得到的第三输出值和第二标签的机器学习,从而生成第三模型,所述表示所述第三图像所包含的被拍摄体的表情肌的动作的产生强度或产生的有无。

2.根据权利要求1所述的机器学习程序,其特征在于,

所述第一图像和所述第二图像的组是相同的被拍摄体的图像的组。

3.根据权利要求1所述的机器学习程序,其特征在于,

所述第一标签基于对所述被拍摄体的表情肌的动作进行测量的测量装置的测量结果而被赋予。

4.根据权利要求1所述的机器学习程序,其特征在于,

所述第一图像和所述第二图像的组是彼此的图像中的所述被拍摄体的表情肌的动作之差为特定值以上的图像的组。

5.根据权利要求1所述的机器学习程序,其特征在于,

生成所述第三模型的处理通过包括基于包含所述第三图像的动态图像的特征量的机器学习,从而生成所述第三模型。

6.根据权利要求5所述的机器学习程序,其特征在于,

所述特征量是将所述动态图像所包含的图像组输入到所述训练完毕模型而得到的输出值的时间序列数据、与该时间序列数据的分布相关的特征量、基于所述时间序列数据的分布从所述图像组选择的一个或者多个图像、以及该一个或者多个图像的特征量的至少一个。

7.一种机器学习方法,其特征在于,

由计算机执行如下处理:

根据包含第一图像和第二图像的组以及第一标签的训练数据的输入,执行第一模型的机器学习,从而生成训练完毕模型,其中,所述第一标签表示所述第一图像和所述第二图像中的哪个图像是被拍摄体的表情肌的动作大的图像,所述第一模型的机器学习基于将所述第一图像输入到第一模型而得到的第一输出值、将所述第二图像输入到与所述第一模型共享参数的第二模型而得到的第二输出值、以及所述第一标签,

通过基于将第三图像输入到所述训练完毕模型而得到的第三输出值和第二标签的机器学习,从而生成第三模型,所述第二标签表示所述第三图像所包含的被拍摄体的表情肌的动作的产生强度或产生的有无。

8.根据权利要求7所述的机器学习方法,其特征在于,

所述第一图像和所述第二图像的组是相同的被拍摄体的图像的组。

9.根据权利要求7所述的机器学习方法,其特征在于,

所述第一标签基于对所述被拍摄体的表情肌的动作进行测量的测量装置的测量结果而被赋予。

10.根据权利要求7所述的机器学习方法,其特征在于,

所述第一图像和所述第二图像的组是彼此的图像中的所述被拍摄体的表情肌的动作之差为特定值以上的图像的组。

11.根据权利要求7所述的机器学习方法,其特征在于,

生成所述第三模型的处理通过包括基于包含所述第三图像的动态图像的特征量的机器学习,从而生成所述第三模型。

12.根据权利要求11所述的机器学习方法,其特征在于,

所述特征量是将所述动态图像所包含的图像组输入到所述训练完毕模型而得到的输出值的时间序列数据、与该时间序列数据的分布相关的特征量、基于所述时间序列数据的分布从所述图像组选择的一个或者多个图像、以及该一个或者多个图像的特征量的至少一个。

13.一种推定装置,具有控制部,

所述控制部执行如下处理:

向第一机器学习模型输入第三图像,获取第一输出结果,所述第一机器学习模型通过基于包含第一图像和第二图像的组以及第一标签的训练数据的机器学习而生成,所述第一标签表示所述第一图像和所述第二图像中的哪个图像是被拍摄体的表情肌的动作大的图像,

向第二机器学习模型输入所述第一输出结果,推定所述第三图像所包含的被拍摄体的表情肌的动作的产生强度或产生的有无,所述第二机器学习模型通过基于包含将第四图像输入到所述机器学习模型而得到的第二输出结果、以及表示所述第四图像所包含的被拍摄体的表情肌的动作的产生强度的第二标签的训练数据的机器学习而生成。

14.根据权利要求13所述的推定装置,其特征在于,

所述第一图像和所述第二图像的组是相同的被拍摄体的图像的组。

15.根据权利要求13所述的推定装置,其特征在于,

所述第一标签基于对所述被拍摄体的表情肌的动作进行测量的测量装置的测量结果而被赋予。

16.根据权利要求13所述的推定装置,其特征在于,

所述第一图像和所述第二图像的组是彼此的图像中的所述被拍摄体的表情肌的动作之差为特定值以上的图像的组。

17.根据权利要求13所述的推定装置,其特征在于,

在所述推定的处理中,向所述第二机器学习模型输入所述第一输出结果和基于包含所述第三图像的动态图像的特征量,所述第二机器学习模型通过包括基于包含所述第四图像的动态图像的特征量的机器学习而生成。

18.根据权利要求17所述的推定装置,其特征在于,

所述特征量是将所述动态图像所包含的图像组输入到所述训练完毕模型而得到的输出值的时间序列数据、与该时间序列数据的分布相关的特征量、基于所述时间序列数据的分布从所述图像组选择的一个或者多个图像以及该一个或者多个图像的特征量的至少一个。

机器学习程序、机器学习方法以及推定装置

技术领域

[0001] 本发明的实施方式涉及表情的推定技术。

背景技术

[0002] 为了推定表情，首先必须规定作为推定对象的表情的描述方法。至今，提出了被称为AU(Action Unit:动作单元)的表情的描述方法。AU表示基于面肌的解剖学的见解定义的与表情露出有关的表情肌的动作。

[0003] 推定AU的AU推定引擎的代表性的方式以基于大量的教师数据的机器学习为基础，使用面部表情的图像数据和各AU的Occurrence(产生的有无)、Intensity(产生强度)，作为教师数据。另外，教师数据的Occurrence、Intensity由被称为Coder(编码器)的专家Annotation(注释)。此外，在以下，有时仅列举Intensity进行说明，但对于Occurrence也同样。

[0004] 专利文献1:美国专利申请公开第2019/294868号说明书;

[0005] 专利文献2:日本特开2020-57111号公报;

[0006] 专利文献3:日本特开2018-36734号公报。

[0007] AU的Intensity作为表情肌的动作被统一定义。但是，能够从外部观察的是皮肤表面的移动量、外表的变化(皱纹等)，这根据年龄、骨骼、肥胖的程度、皮肤和表情肌的关联方式等存在各种变更，因此难以对所有人定义统一的Intensity的边界基准(以下，也有时仅称为“基准”)。因此，能够从外部观察的Intensity的边界基准不得不变得模糊。

[0008] 编码器例如一边观察作为正解标签赋予Intensity的对象的被拍摄体的动态图像，一边推测表情肌的动作并对一个一个帧图像赋予Intensity。然而，能够从外部观察的Intensity的边界基准模糊，对所有人不是统一的基准，因此编码器赋予的正解标签根据被拍摄体，边界的基准有时偏移。

[0009] 因此，在上述的现有技术中，有时生成包含这样的基准的不同的数据进行机器学习而捕捉到错误的特征的模型，存在推定精度降低的问题。

发明内容

[0010] 在一个方面，其目的在于提供能够提高表情推定的精度的机器学习程序、机器学习方法以及推定装置。

[0011] 在第一方案中，机器学习程序使计算机执行生成训练完毕模型的处理和生成第三模型的处理。在生成训练完毕模型的处理中，根据包含第一图像和第二图像的组以及第一标签的训练数据的输入，执行第一模型的机器学习，从而生成训练完毕模型，所述第一标签表示第一图像和第二图像中的哪个图像是被拍摄体的表情肌的动作大的图像，所述第一模型的机器学习基于将第一图像输入到第一模型而得到的第一输出值、将第二图像输入到与第一模型共享参数的第二模型而得到的第二输出值、以及第一标签。在生成第三模型的处理中，通过基于将第三图像输入到训练完毕模型而得到的第三输出值和第二标签的机器学

习,从而生成第三模型,所述第二标签表示第三图像所包含的被拍摄体的表情肌的动作的产生强度或产生的有无。

[0012] 能够提高表情推定的精度。

附图说明

[0013] 图1是说明实施方式的概要的说明图。

[0014] 图2是表示第一实施方式所涉及的信息处理装置的功能结构例的框图。

[0015] 图3是表示第一实施方式所涉及的信息处理装置的动作例的流程图。

[0016] 图4是表示第一实施方式所涉及的信息处理装置的动作例的流程图。

[0017] 图5是表示第二实施方式所涉及的信息处理装置的功能结构例的框图。

[0018] 图6是表示第三实施方式所涉及的信息处理装置的功能结构例的框图。

[0019] 图7是表示第四实施方式所涉及的信息处理装置的功能结构例的框图。

[0020] 图8是说明第四实施方式所涉及的信息处理装置的处理内容的说明图。

[0021] 图9是表示计算机结构的一个例子的框图。

具体实施方式

[0022] 以下,参照附图,对实施方式所涉及的机器学习程序、机器学习方法以及推定装置进行说明。对实施方式中具有相同的功能的结构标注相同的附图标记,省略重复的说明。此外,在以下的实施方式中说明的机器学习程序、机器学习方法以及推定装置只不过表示一个例子,并不限定实施方式。另外,以下的各实施方式也可以在不矛盾的范围内适当地组合。

[0023] (概要)

[0024] 在以下进行说明的第一、第二、第三以及第四实施方式(汇总而言本实施方式)中,在AU推定引擎的机器学习中,从学习用的学习数据集(训练数据)中,对于Intensity的边界基准,排除受到变化的影响的信息进行学习。

[0025] 例如,即使根据年龄、骨骼、肥胖的程度、皮肤和表情肌的关联方式等基准在训练数据所包含的被拍摄体间变化,在同一被拍摄体中基准也不会变化。因此,在本实施方式中,学习训练数据中的同一被拍摄体中的Intensity的顺序关系(哪个Intensity大),生成能够推定保证Intensity的顺序关系的Intensity(称为顺序Intensity)的模型。

[0026] 图1是说明实施方式的概要的说明图。如图1所示,在本实施方式中,将同一被拍摄体a的图像的组即图像对(a_1、a_2)、和表示该图像对中的哪个图像是被拍摄体a的表情肌的动作大的图像的正解标签设为学习数据集D1。

[0027] 例如,在本实施方式中,编码器将作为正解标志而赋予到各图像的Intensity进行比较而得到正解标签。作为一个例子,在图像a_1中的Intensity大于图像a_2中的Intensity的情况下,将“1”设为正解标签。另外,在图像a_1中的Intensity小于图像a_2中的Intensity的情况下,将“0”设为正解标签。在本实施方式中,针对多个被拍摄体的每一个准备该学习数据集D1。

[0028] 接着,在本实施方式中,使用所准备的学习数据集D1,进行神经网络NN的学习,以使能够计算保证顺序关系的顺序Intensity。具体而言,神经网络NN是使用通过一般物体识

别的数据集等预先学习的预先学习模型M1、M2(共享彼此的参数(权重))、评价预先学习模型M1、M2的输出值的loss函数F1的网络。

[0029] 例如,从预先学习模型M1相对于学习数据集D1中的图像a_1的输入,获得顺序Intensity(I_{a1})作为输出值。另外,从预先学习模型M2相对于学习数据集D1中的图像a_2的输入,获得顺序Intensity(I_{a2})作为输出值。loss函数F1根据正解标签L1和预先学习模型M1、M2的输出值(I_{a1},I_{a2}),如果正解标签为1且$I_{a1}>I_{a2}$、或者正解标签为0且$I_{a1}<I_{a2}$则顺序正确,因此将loss值计算得较小。另外,如果正解标签为1且$I_{a1}<I_{a2}$、或者正解标签为0且$I_{a1}>I_{a2}$则顺序不正确,因此loss函数F1将loss值计算得较大,输出计算出的loss值。

[0030] 在本实施方式中,使神经网络NN中的参数(预先学习模型M1的参数)再学习,以使loss函数F1输出的loss值变小。由此,在使用再学习后的参数的模型(训练完毕模型)中,能够对所输入的图像计算保证顺序关系的顺序Intensity。

[0031] 接着,在本实施方式中,根据所生成的模型推定的顺序Intensity和训练数据所包含的编码器赋予的正解标志(例如范围为0～5的Intensity的值),学习将顺序Intensity转换为Intensity的转换函数。该转换函数也可以根据编码器赋予的与Occurrence相关的正解标志,将顺序Intensity转换为Occurrence。

[0032] 此外,由于基准的变化,在顺序Intensity和编码器的正解标志(例如Intensity的值)中包含噪声,但对于转换函数,取单纯的单调递增函数的形式,因此能够不受噪声的影响地进行学习。

[0033] 这样,在本实施方式中,将机器学习分为两个阶段并生成模型以及转换函数,从而对于Intensity的边界基准,排除受到变化的影响的信息进行学习。通过这样的机器学习,在本实施方式中,能够生成正确地捕捉表情肌的动作等与表情推定相关的特征的模型,能够提高AU推定引擎中的表情推定的精度。

[0034] (第一实施方式)

[0035] 图2是表示第一实施方式所涉及的信息处理装置的功能结构例的框图。如图2所示,信息处理装置1具有:成对数据集制成部10、顺序得分学习部11、转换数据集制成部12、转换函数学习部13、顺序得分推定部30以及转换处理部31。

[0036] 在这里,成对数据集制成部10、顺序得分学习部11、转换数据集制成部12以及转换函数学习部13是主要进行执行机器学习的与学习情况(S1)相关的处理的功能部。另外,顺序得分推定部30以及转换处理部31是主要进行使用由机器学习生成的模型从推定对象的图像50推定标签51的与推定情况(S2)相关的处理的功能部。

[0037] 成对数据集制成部10是根据储存用于进行机器学习的训练数据的图像及标签DB40,制成包含同一被拍摄体a的图像对(a_1,a_2)、和表示该图像对的哪个图像是被拍摄体a的表情肌的动作大的图像的正解标签的学习数据集D1的处理部。

[0038] 具体而言,图像及标签DB40例如储存赋予了基于编码器等的正解标志(例如Intensity)和表示被拍摄体a的被拍摄体信息等的元数据的与多个被拍摄体a相关的各图像作为训练数据。此外,在图像及标签DB40储存的训练数据中也可以包含将各图像设为一帧的动态图像。成对数据集制成部10例如根据图像及标签DB40的各图像中的元数据获得同一被拍摄体a的图像的组即图像对。另外,成对数据集制成部10也可以从图像及标签DB40获得彼此的图像中的被拍摄体a的表情肌的动作之差(例如Intensity之差)为特定值以上的

图像对。

[0039] 此外，对于同一被拍摄体a，在本实施方式中，例如被拍摄体信息中的被拍摄体ID是相同的同一人物，但在被拍摄体间Intensity的边界基准不变化的属性(例如年龄、性别、种族)也可以是同一被拍摄体a。另外，同一人物的动态图像被分为多个，以所分的动态图像为单位，在编码器赋予正解标志的情况下，由于Intensity的边界基准的模糊，即使是同一人物，也有可能由于动态图像而基准偏离，因此在这样的情况下，也可以仅在动态图像相同的情况下，作为同一被拍摄体a进行处理。

[0040] 接着，成对数据集制成部10通过比较图像对的元数据所包含的正解标志，获得表示图像对的哪个图像是被拍摄体a的表情肌的动作大的图像的正解标签。通过反复上述的处理，在成对数据集制成部10中，制成关于多个被拍摄体a的每一个的学习数据集D1。

[0041] 顺序得分学习部11是使用成对数据集制成部10制成的学习数据集D1，进行神经网络NN的学习，以使能够计算保证顺序关系的顺序Intensity的处理部。

[0042] 具体而言，顺序得分学习部11通过从储存预先学习模型M1、M2的参数(权重)的预先学习模型DB41获取到的参数构建预先学习模型M1、M2(共享彼此的参数)。预先学习模型M1、M2例如是利用ImageNet数据集预先学习的VGG16等，将最终层的输出作为一维。

[0043] 另外，顺序得分学习部11根据预先学习模型M1、M2的输出值和正解标签L1，通过loss函数F1，得到loss值。在这里，关于loss函数F1，例如，使用下式(1)那样的函数。

[0044] [式1]

[0045] $$loss = \sum_{(a_1,a_2)} \max\left(0, m - L_{a_1,a_2}\cdot\left(I_{a_1} - I_{a_2}\right) - \left(1 - L_{a_1,a_2}\right)\cdot\left(I_{a_2} - I_{a_1}\right)\right) \quad \cdots(1)$$

[0046] $$\begin{cases} L_{a_1,a_2}: \text{学习数据集的正解标签}\ (0\ or\ 1) \\ \cdot \text{图像}a_1\text{の}\mathrm{Intensity} < \text{图像}a_2\text{の}\mathrm{Intensity} \Rightarrow L_{a_1,a_2} = 0 \\ \cdot \text{图像}a_1\text{の}\mathrm{Intensity} > \text{图像}a_2\text{の}\mathrm{Intensity} \Rightarrow L_{a_1,a_2} = 1 \\ m = 1 \end{cases}$$

[0047] 接着，顺序得分学习部11使神经网络NN中的参数(预先学习模型M1的参数)再学习，以使loss函数F1输出的loss值变小。此外，对于神经网络NN，作为其它例子，也可以为ResNet。另外，也可以利用使用面部图像数据集进行预先学习的模型。并且，也可以利用不进行预先学习而通过随机数使权重初始化的模型。顺序得分学习部11将通过再学习而得到的参数储存于学习模型DB20。

[0048] 转换数据集制成部12是制成用于转换函数学习部13学习转换函数的数据集的处理部。具体而言，转换数据集制成部12读取图像及标签DB40的训练数据所包含的图像和赋予到图像的正解标志(例如Intensity或者Occurrence)。接着，转换数据集制成部12将读取到的图像输入到顺序得分推定部30，并获取来自通过顺序得分学习部11的学习生成的训练完毕模型的输出(顺序Intensity)。接着，转换数据集制成部12制成对于获取到的顺序Intensity将赋予到图像的正解标志的Intensity(或者Occurrence)作为正解标签的数据集。

[0049] 转换函数学习部13是使用转换数据集制成部12制成的数据集，对将顺序Intensity转换为Intensity(或者Occurrence)的转换函数进行机器学习的处理部。转换函

数学习部13将通过机器学习获得的与转换函数相关的参数储存于转换模型DB21。

[0050] 具体而言,转换函数学习部13根据转换数据集制成部12制成的数据集,进行回归学习求出转换函数,以使输出范围为0～5的Intensity的值作为连续值。另外,转换函数学习部13也可以进行分类学习求出转换函数,以使输出[0、1、2、3、4、5]的离散值(在Occurrence的情况下,0、1的离散值)。在回归学习或者分类学习的任一种情况下,转换函数学习部13都能够通过神经网络、SVM(Support Vector Machine:支持向量机)等公知的机器学习方法获得转换函数。

[0051] 顺序得分推定部30是根据所输入的图像(例如作为推定对象输入的图像50)推定顺序Intensity的处理部。

[0052] 具体而言,顺序得分推定部30从学习模型DB20读出顺序得分学习部11进行学习而生成的模型的参数(权重),构建模型。接着,顺序得分推定部30将推定对象的图像50输入到模型,从而推定顺序Intensity。

[0053] 转换处理部31是通过转换函数将顺序得分推定部30推定出的顺序Intensity转换为Intensity(或者Occurrence)的处理部。转换处理部31将转换后的Intensity(或者Occurrence)作为表示对于图像50的推定结果的图像50进行输出。

[0054] 具体而言,转换处理部31从转换模型DB21读出转换函数学习部13进行学习后的转换函数的参数,构建转换函数。接着,转换处理部31通过所构建的转换函数,将顺序得分推定部30推定出的顺序Intensity转换为Intensity(或者Occurrence),并作为图像50进行输出。

[0055] 此外,转换处理部31在对转换函数进行回归学习以输出连续值的情况下,将输出范围限制在定义域(0～5)内,因此也可以修正为以上限5、下限0进行限制。另外,转换处理部31也可以通过小数点以下的四舍五入,将输出值离散化为0、1、2、3、4、5这六个阶段。

[0056] 图3是表示第一实施方式所涉及的信息处理装置1的动作例的流程图。更具体而言,图3是表示信息处理装置1中的学习情况(S1)的动作例的流程图。

[0057] 如图3所示,若开始处理,则成对数据集制成部10根据图像及标签DB40制成包含与同一被拍摄体a相关的图像对和表示Intensity的顺序关系的正解标签的学习数据集D1(S11)。

[0058] 接着,顺序得分学习部11根据所制成的学习数据集D1进行神经网络NN的再学习(S12),将所学习的神经网络NN的参数(预先学习模型M1的参数)储存于学习模型DB20(S13)。接着,转换数据集制成部12从图像及标签DB40读入图像和赋予到图像的正解标志(例如Intensity)(S14)。

[0059] 接着,转换数据集制成部12将读取到的图像输入到顺序得分推定部30,并获取来自通过顺序得分学习部11的学习而生成的训练完毕模型的输出(顺序Intensity)。接着,转换数据集制成部12制成对于获取到的顺序Intensity将赋予到图像的正解标志的Intensity作为正解标签的学习数据集(S15)。

[0060] 接着,转换函数学习部13通过转换数据集制成部12制成的学习数据集,进行转换函数的学习(S16)。接着,转换函数学习部13将所学习的转换函数的参数储存于转换模型DB21(S17),结束处理。

[0061] 图4是表示第一实施方式所涉及的信息处理装置1的动作例的流程图。更具体而

言，图4是表示信息处理装置1中的推定情况(S2)的动作例的流程图。

[0062] 如图4所示，若开始处理，则顺序得分推定部30获取推定对象的图像50(S21)。接着，顺序得分推定部30从学习模型DB20获取参数并构建神经网络(训练完毕模型)(S22)。

[0063] 接着，顺序得分推定部30向所构建的神经网络(训练完毕模型)输入图像50并得到输出值，从而推定相对于图像50的顺序Intensity(S23)。

[0064] 接着，转换处理部31从转换模型DB21获取参数并构建转换函数(S24)。接着，转换处理部31通过所构建的转换函数，将顺序得分推定部30推定出的顺序Intensity转换为Intensity(S25)。接着，转换处理部31将转换后的Intensity作为推定出的标签51进行输出(S26)。

[0065] (第二实施方式)

[0066] 在第二实施方式中，关于训练数据的被赋予的正解标志，不仅包括编码器的正解标志，还包括对被拍摄体的表情肌的动作进行测量的测量装置的测量结果(Intensity或表情肌移动量)。

[0067] 图5是表示第二实施方式所涉及的信息处理装置的功能结构例的框图。如图5所示，图像及标签DB40a储存赋予了测量装置的测量结果(Intensity或表情肌移动量)的正解标志和表示被拍摄体a的被拍摄体信息等元数据的与多个被拍摄体a相关的各图像作为训练数据。

[0068] 信息处理装置1a的成对数据集制成部10a参照图像及标签DB40、40a，制成包含同一被拍摄体a的图像对(a_1，a_2)、和表示该图像对的哪个图像是被拍摄体a的表情肌的动作大的图像的正解标签的学习数据集D1。

[0069] 在编码器赋予的正解标志(例如Intensity)中，由于人为错误，存在即使是同一被拍摄体a也包含噪声的情况。与此相对，在测量装置的测量结果中，不会产生基于人为错误的噪声。因此，将测量装置的测量结果包含于正解标志中，从而能够期待提高顺序关系的学习精度的效果。

[0070] (第三实施方式)

[0071] 在第三实施方式中，不使用编码器赋予正解标志的训练数据，而使用赋予测量装置的测量结果作为正解标志的训练数据进行第一阶段的学习。此外，对于第二阶段的学习(转换函数的学习)，使用编码器赋予正解标志的训练数据。

[0072] 图6是表示第三实施方式所涉及的信息处理装置的功能结构例的框图。如图6所示，信息处理装置1b的成对数据集制成部10a不参照图像及标签DB40，而参照图像及标签DB40a制成学习数据集D1。此外，转换数据集制成部12参照图像及标签DB40制成用于转换函数学习部13学习转换函数的数据集。

[0073] 这样，在信息处理装置1b中，不使用编码器赋予的正解标志，而使用测量装置的测量结果学习顺序关系，从而能够期待进一步提高顺序关系的学习精度的效果。

[0074] (第四实施方式)

[0075] 在第四实施方式中，在学习情况(S1)中，通过包括包含图像作为一帧图像的动态图像的特征量的机器学习对转换函数进行机器学习。另外，在推定情况(S2)中，根据包含推定对象的图像50作为一帧图像的元动态图像的特征量，通过转换函数推定标签51。

[0076] 图7是表示第四实施方式所涉及的信息处理装置的功能结构例的框图。如图7所

示,信息处理装置1c的转换数据集制成部12a与图像及标签DB40所包含的图像一起,获取将该图像设为一帧的动态图像。接着,转换数据集制成部12a通过分析获取到的动态图像,获取动态图像的特征量,将获取到的特征量包含于用于学习转换函数的数据集中。

[0077] 信息处理装置1c的转换函数学习部13a使用转换数据集制成部12制成的数据集,对将顺序Intensity转换为Intensity(或者Occurrence)的转换函数进行机器学习时,进行包含基于动态图像的特征量的机器学习。

[0078] 图8是说明第四实施方式所涉及的信息处理装置1c的处理内容的说明图。图8中的S30表示转换数据集制成部12a执行的具体的处理。另外,图8中的S40表示与转换函数学习部13a中的处理有关的结构。

[0079] 如图8所示,转换数据集制成部12a将图像及标签DB40的图像42输入到顺序得分推定部30,获取来自训练完毕模型(神经网络(VGG))的输出值(顺序Intensity)(S31)。

[0080] 另外,转换数据集制成部12a将使图像42为一帧的动态图像43输入到顺序得分推定部30,从训练完毕模型(神经网络(VGG))获取时间序列的输出值(顺序Intensity的时间序列数据)(S32)。

[0081] 接着S32,转换数据集制成部12a提取与时间序列数据的分布相关的特征量(S33)。具体而言,转换数据集制成部12a求出基于时间序列数据的直方图(例如顺序Intensity－10以下的频率、－10～－9的频率、－9～－8的频率、…、9～10的频率、10以上的频率)。另外,转换数据集制成部12a求出基于时间序列数据的百分位数(0－th percentile、10－th percentile、…、100－th percentile)。

[0082] 另外,接着S32,转换数据集制成部12a进行基于时间序列数据的图像选定处理(S34)。具体而言,转换数据集制成部12a在时间序列数据中选定顺序Intensity满足规定条件的一个或者多个图像(例如最小的顺序Intensity的图像的选定)。

[0083] 接着S34,转换数据集制成部12a对于选定的图像,进行图像特征量、面部特征量等的提取处理(S35)。具体而言,转换数据集制成部12a提取图像的SIFT特征量和该图像的地标等。

[0084] 在转换函数学习部13a中,将通过S30制成的数据集输入到作为转换函数的模型的LSTM22、VGG23以及全结合神经网络24,得到Intensity25。

[0085] 具体而言,构成为向全结合神经网络24输入基于S31的图像的顺序Intensity和其它的特征量。例如,在被输入时间序列数据本身的网络中使用LSTM22等的面向时间序列数据的网络。另外,在被输入图像数据本身的网络中使用VGG23等的面向图像数据的网络。另外,构成为LSTM22的输出以及VGG23的输出与全结合神经网络24连接。

[0086] 返回到图7,转换处理部31a通过根据转换模型DB21的参数构建的转换函数,根据顺序得分推定部30推定出的顺序Intensity和元动态图像50a的特征量推定标签51。具体而言,转换处理部31a与转换数据集制成部12a同样地求出元动态图像50a的特征量,与顺序Intensity一起输入到转换函数来推定标签51。

[0087] 这样,在第四实施方式中,通过包含动态图像的特征量,能够实现更高精度的表情推定。

[0088] 如以上那样,信息处理装置1制成包括图像及标签DB40所包含的图像对、和表示图像对中的哪个图像是被拍摄体的表情肌的动作大的图像的正解标签的成对数据集。接着,

信息处理装置1通过执行基于将图像对中的第一图像输入到预先学习模型M1而得到的输出值、将图像对中的第二图像输入到与预先学习模型M1共享参数的预先学习模型M2而得到的输出值以及第一标签的预先学习模型M1的机器学习，生成训练完毕模型。接着，信息处理装置1通过基于将图像及标签DB40所包含的第三图像输入到训练完毕模型而得到的输出值和表示该图像所包含的被拍摄体的表情肌的动作的产生强度或产生的有无的标签的机器学习生成模型(转换函数)。

[0089] 这样，通过分为两个阶段的机器学习，信息处理装置1能够生成用于正确地捕捉表情肌的动作等与表情推定相关的特征并进行推定的模型。另外，信息处理装置1使用这样生成的模型，进行相对于推定对象的图像50的标签51的推定、即表情推定，从而能够提高表情推定的精度。

[0090] 另外，与训练完毕模型的生成相关的图像对是相同的被拍摄体的图像对。例如，即使根据年龄、骨骼、肥胖的程度、皮肤和表情肌的关联方式等而基准在图像及标签DB40所包含的图像的被拍摄体间变化，基准在相同的被拍摄体中也不会变化。因此，通过相同的被拍摄体的图像对进行训练完毕模型的生成，从而在信息处理装置1中，能够生成能够更适当地推定被拍摄体的表情肌的动作是否大(Intensity的顺序关系)的训练完毕模型。

[0091] 另外，成对数据集中的正解标签基于对被拍摄体的表情肌的动作进行测量的测量装置的测量结果(例如Intensity)被赋予。例如，即使是相同的表情肌的动作，其移动量也根据人而有所不同，与作为专家的编码器赋予的情况同样，在测量装置的测量结果中也根据被拍摄体而基准偏移。在信息处理装置1中，在赋予基于这样的测量装置的测量结果的正解标签的情况下，也能够生成用于正确地捕捉与表情推定相关的特征并进行推定的模型。

[0092] 另外，与训练完毕模型的生成相关的图像对彼此的图像中的被拍摄体的表情肌的动作之差为特定值以上。这样，使用被拍摄体的表情肌的动作之差为特定值以上，被拍摄体的表情肌的动作有明确的差异的图像对，从而能够生成更高精度的模型。

[0093] 另外，与训练完毕模型的生成相关的图像对也可以是表情肌的动作的大小相同的对。该情况下，作为loss函数F1，例如使用下式(2)。

[0094] [式2]

[0095] $$loss = \sum_{(a_1,a_2)} |L_{a_1,a_2}| * \max\left(0, m - L_{a_1,a_2} \cdot (l_{a_1} - l_{a_2})\right) + (1 - |L_{a_1,a_2}|) * |l_{a_1} - l_{a_2}| \quad \cdots (2)$$

[0096] $$\begin{pmatrix} L_{a_1,a_2}: \text{学习数据集的正解标签}(-1\ or\ 0\ or\ 1) \\ \cdot \text{图像}a_1\text{の}\mathrm{Intensity} < \text{图像}a_2\text{の}\mathrm{Intensity} \Rightarrow L_{a_1,a_2} = -1 \\ \cdot \text{图像}a_1\text{の}\mathrm{Intensity} = \text{图像}a_2\text{の}\mathrm{Intensity} \Rightarrow L_{a_1,a_2} = 0 \\ \cdot \text{图像}a_1\text{の}\mathrm{Intensity} > \text{图像}a_2\text{の}\mathrm{Intensity} \Rightarrow L_{a_1,a_2} = 1 \\ m = 1 \end{pmatrix}$$

[0097] 在信息处理装置1中，也使用表情肌的动作的大小相同的对，从而能够生成更高精度的模型。

[0098] 另外，信息处理装置通过包括基于包含第三图像的动态图像的特征量的机器学习生成模型(转换函数)。这样，包含动态图像的特征量，从而信息处理装置能够生成更高精度的模型。

[0099] 另外,基于包含第三图像的动态图像的特征量也可以是基于动态图像所包含的图像组的时间序列数据、与该时间序列数据的分布相关的特征量、基于时间序列数据的分布从图像组选择的一个或者多个图像、以及该一个或者多个图像的特征量的至少一个。在信息处理装置中,包含这样的特征量进行机器学习,从而能够生成更高精度的模型。

[0100] (其它)

[0101] 此外,图示的各装置的各构成要素不一定需要物理上如图示那样构成。即,各装置的分散及综合的具体形态并不限于图示的内容,能够根据各种负荷、使用状况等,以任意单位在功能上或者物理上分散及综合其全部或者一部分来构成。例如,信息处理装置1、1a、1b、1c中的进行学习情况(S1)的功能结构和进行推定情况(S2)的功能结构可以是分离的结构,也可以通过分别独立的装置结构来实现。

[0102] 另外,在信息处理装置1、1a、1b、1c中进行的成对数据集制成部10、10a、顺序得分学习部11、转换数据集制成部12、12a、转换函数学习部13、13a、顺序得分推定部30以及转换处理部31、31a的各种处理功能也可以在CPU(或者MPU、MCU(Micro Controller Unit:微控制器)等微型计算机、GPU(Graphics Processing Unit:图形处理器))上,执行其全部或者任意的一部分。另外,当然各种处理功能也可以在通过CPU(或者MPU、MCU等微型计算机、GPU)分析执行的程序上、或者基于布线逻辑的硬件上,执行其全部或者任意的一部分。另外,在信息处理装置1、1a、1b、1c中进行的各种处理功能也可以通过云计算,由多个计算机协作来执行。

[0103] (计算机构成例)

[0104] 然而,在上述的实施方式中说明的各种处理能够通过由计算机执行预先准备的程序来实现。因此,在以下,对执行具有与上述的实施方式相同的功能的程序的计算机结构(硬件)的一个例子进行说明。图9是表示计算机结构的一个例子的框图。

[0105] 如图9所示,计算机200具有执行各种运算处理的CPU201、接受数据输入的输入装置202、监视器203、扬声器204。另外,计算机200具有从存储介质读取程序等的介质读取装置205、用于与各种装置连接的接口装置206、用于通过有线或者无线与外部设备通信连接的通信装置207。另外,计算机200具有暂时存储各种信息的RAM208、硬盘装置209。另外,计算机200内的各部(201～209)与总线210连接。

[0106] 在硬盘装置209中存储有用于执行在上述的各实施方式中说明的功能结构(例如成对数据集制成部10、10a、顺序得分学习部11、转换数据集制成部12、12a、转换函数学习部13、13a、顺序得分推定部30以及转换处理部31、31a)中的各种的处理的程序211。另外,在硬盘装置209中存储有程序211参照的各种数据212。输入装置202例如从操作者接受操作信息的输入。监视器203例如显示操作者进行操作的各种画面。接口装置206例如连接有印刷装置等。通信装置207与LAN(Local Area Network:局域网)等通信网络连接,在与经由通信网络的外部设备之间交换各种信息。

[0107] CPU201读出存储于硬盘装置209的程序211,在RAM208中展开并执行,从而进行与上述的功能结构(例如成对数据集制成部10、10a、顺序得分学习部11、转换数据集制成部12、12a、转换函数学习部13、13a、顺序得分推定部30以及转换处理部31、31a)相关的各种处理。此外,程序211也可以不存储于硬盘装置209。例如,也可以读出存储于计算机200可读取的存储介质的程序211并执行。计算机200可读取的存储介质例如对应有CD-ROM、DVD盘、USB

(Universal Serial Bus:通用串行总线)存储器等便携式记录介质、闪存等半导体存储器、硬盘驱动器等。另外,也可以在与公用线路、互联网、LAN等连接的装置中存储该程序211,计算机200从这些中读出程序211并执行。

[0108] 附图标记说明

[0109] 1、1a、1b、1c…信息处理装置;10、10a…成对数据集制成部;11…顺序得分学习部;12、12a…转换数据集制成部;13、13a…转换函数学习部;20…学习模型DB;21…转换模型DB;22…LSTM;23…VGG;24…全结合神经网络;25…Intensity;30…顺序得分推定部;31、31a…转换处理部;40、40a…图像及标签DB;41…预先学习模型DB;42…图像;43…动态图像;50…图像;50a…元动态图像;51…标签;200…计算机;201…CPU;202…输入装置;203…监视器;204…扬声器;205…介质读取装置;206…接口装置;207…通信装置;208…RAM;209…硬盘装置;210…总线;211…程序;212…各种数据;a…被拍摄体;D1…学习数据集;F1…loss函数;L1…正解标签;M1、M2…预先学习模型;NN…神经网络。

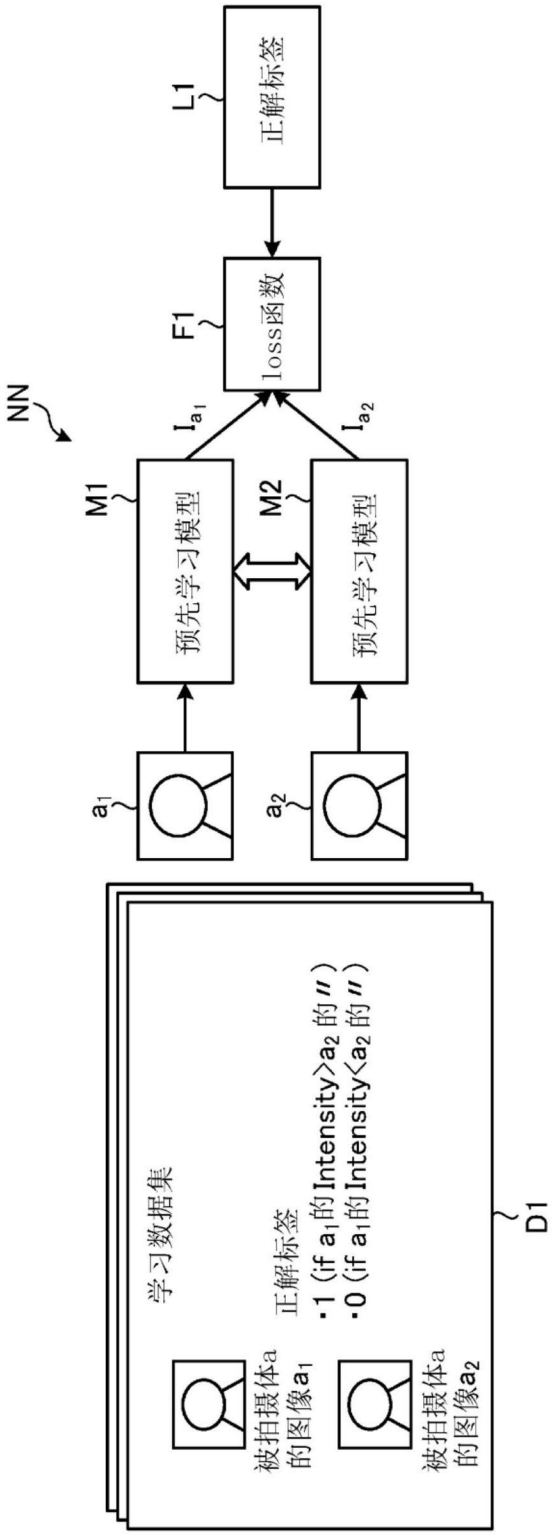

L1
正解标签
F1
loss函数
NN
l_{a_1}
l_{a_2}
M1
预先学习模型
M2
预先学习模型
a_1
a_2
学习数据集
正解标签
·1 (if a_1的Intensity>a_2的〃)
·0 (if a_1的Intensity<a_2的〃)
被拍摄体a的图像a_1
被拍摄体a的图像a_2
D1

图1

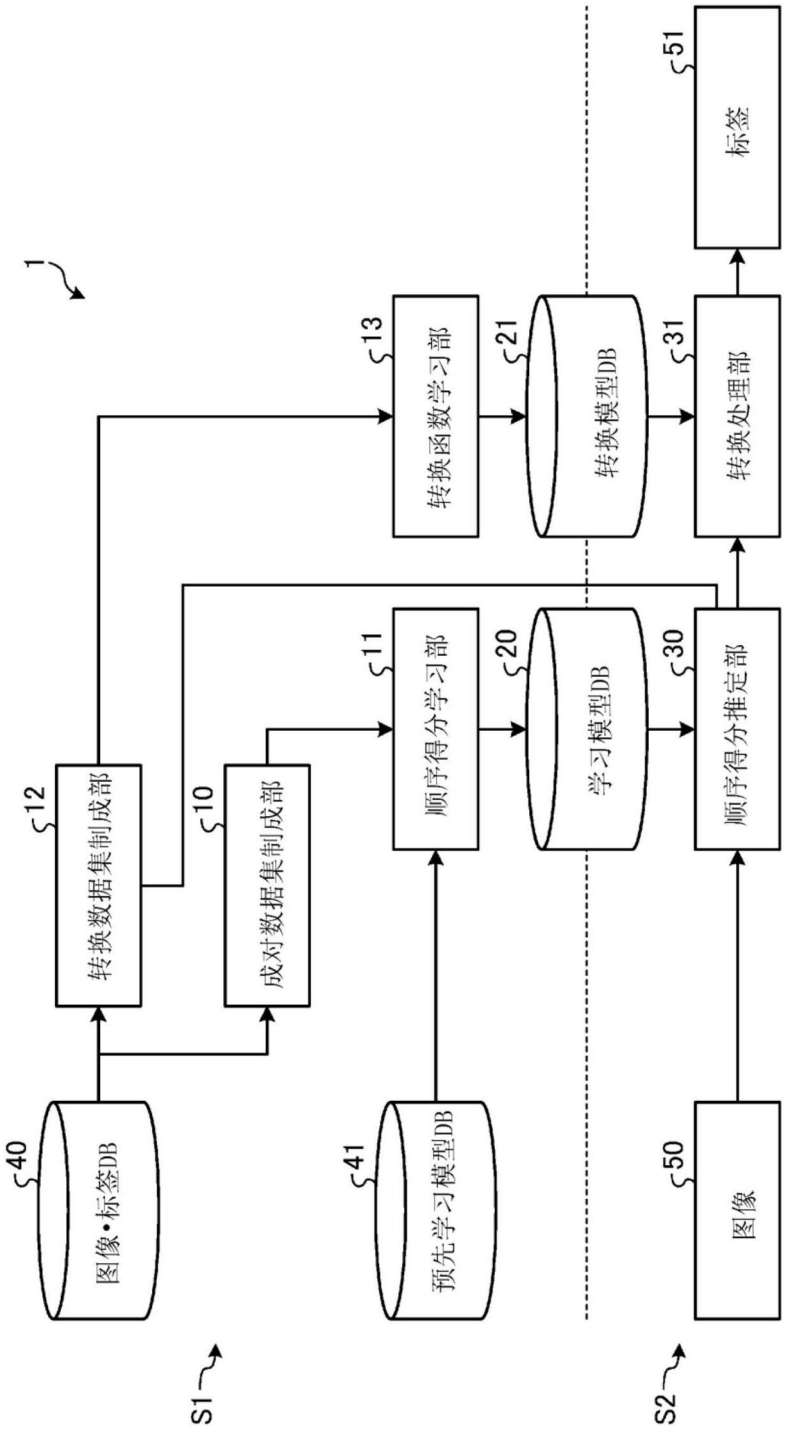

1
40
图像·标签DB
41
预先学习模型DB
50
图像
51
标签
12
转换数据集制成部
10
成对数据集制成部
11
顺序得分学习部
13
转换函数学习部
20
学习模型DB
21
转换模型DB
30
顺序得分推定部
31
转换处理部
S1
S2

图2

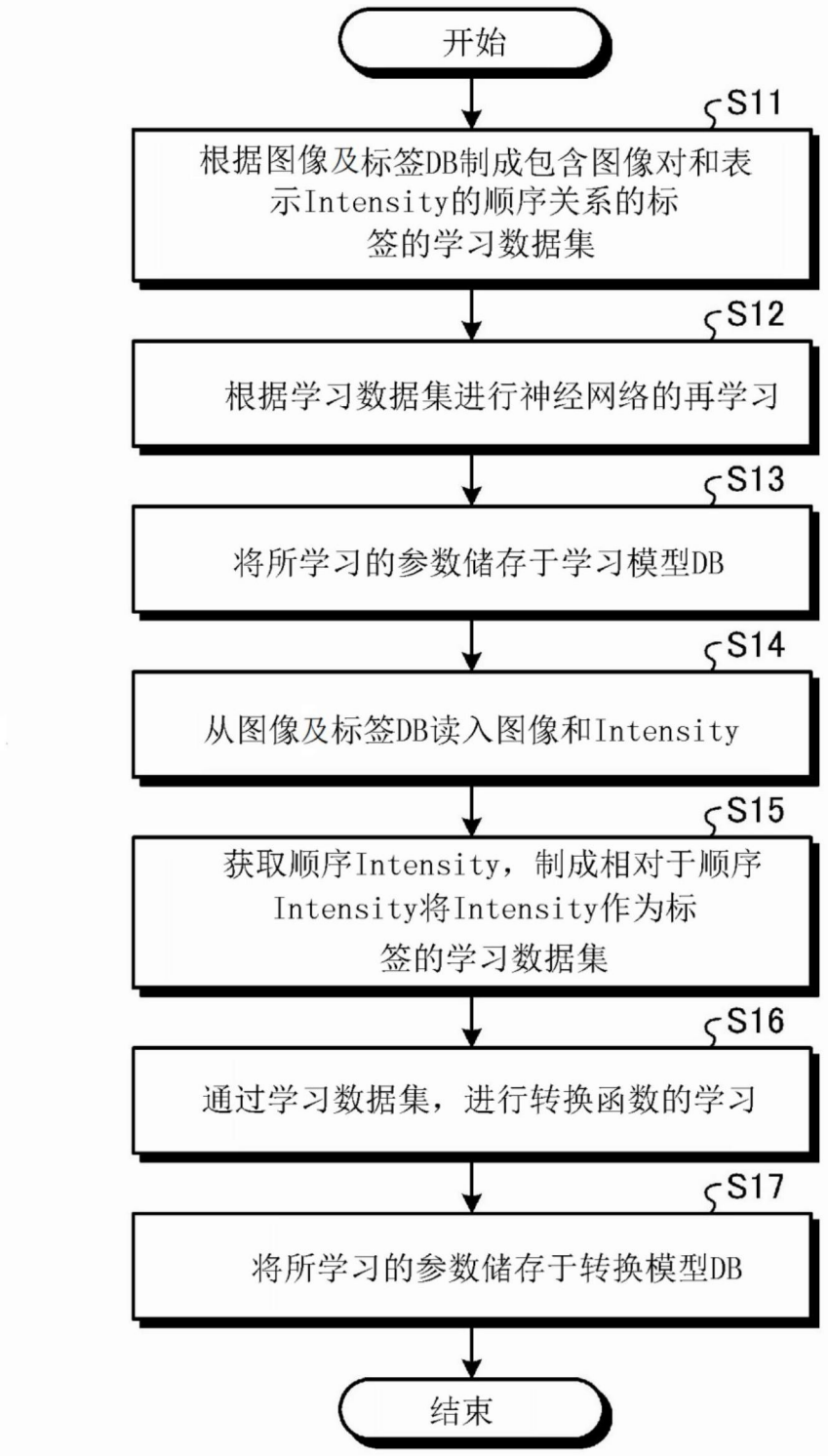

开始
S11
根据图像及标签DB制成包含图像对和表示Intensity的顺序关系的标签的学习数据集
S12
根据学习数据集进行神经网络的再学习
S13
将所学习的参数储存于学习模型DB
S14
从图像及标签DB读入图像和Intensity
S15
获取顺序Intensity，制成相对于顺序Intensity将Intensity作为标签的学习数据集
S16
通过学习数据集，进行转换函数的学习
S17
将所学习的参数储存于转换模型DB
结束

图3

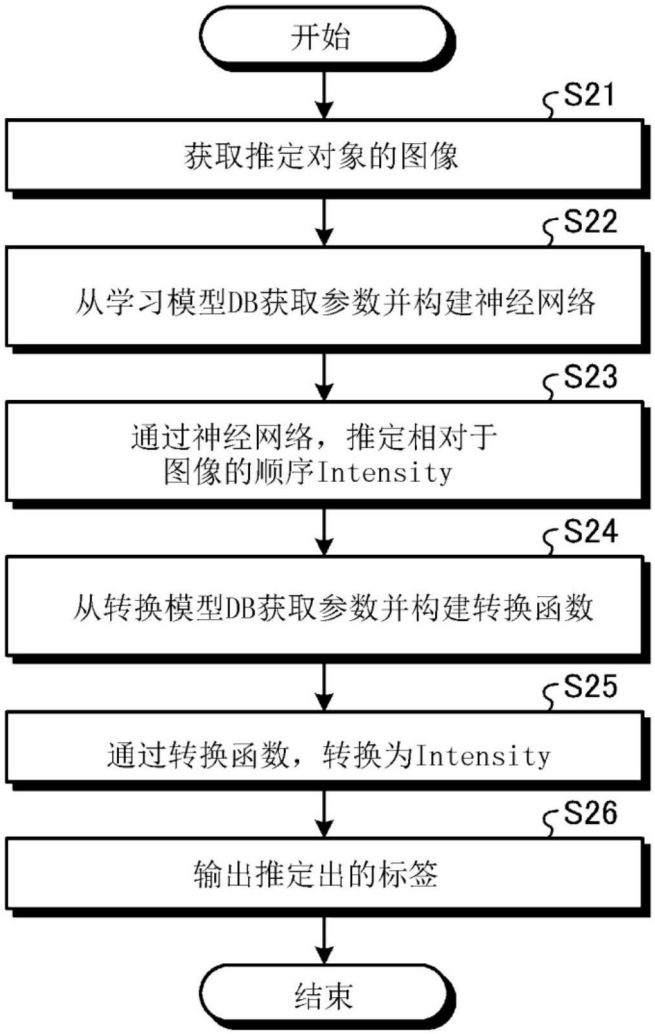

开始
S21
获取推定对象的图像
S22
从学习模型DB获取参数并构建神经网络
S23
通过神经网络，推定相对于
图像的顺序Intensity
S24
从转换模型DB获取参数并构建转换函数
S25
通过转换函数，转换为Intensity
S26
输出推定出的标签
结束

图4

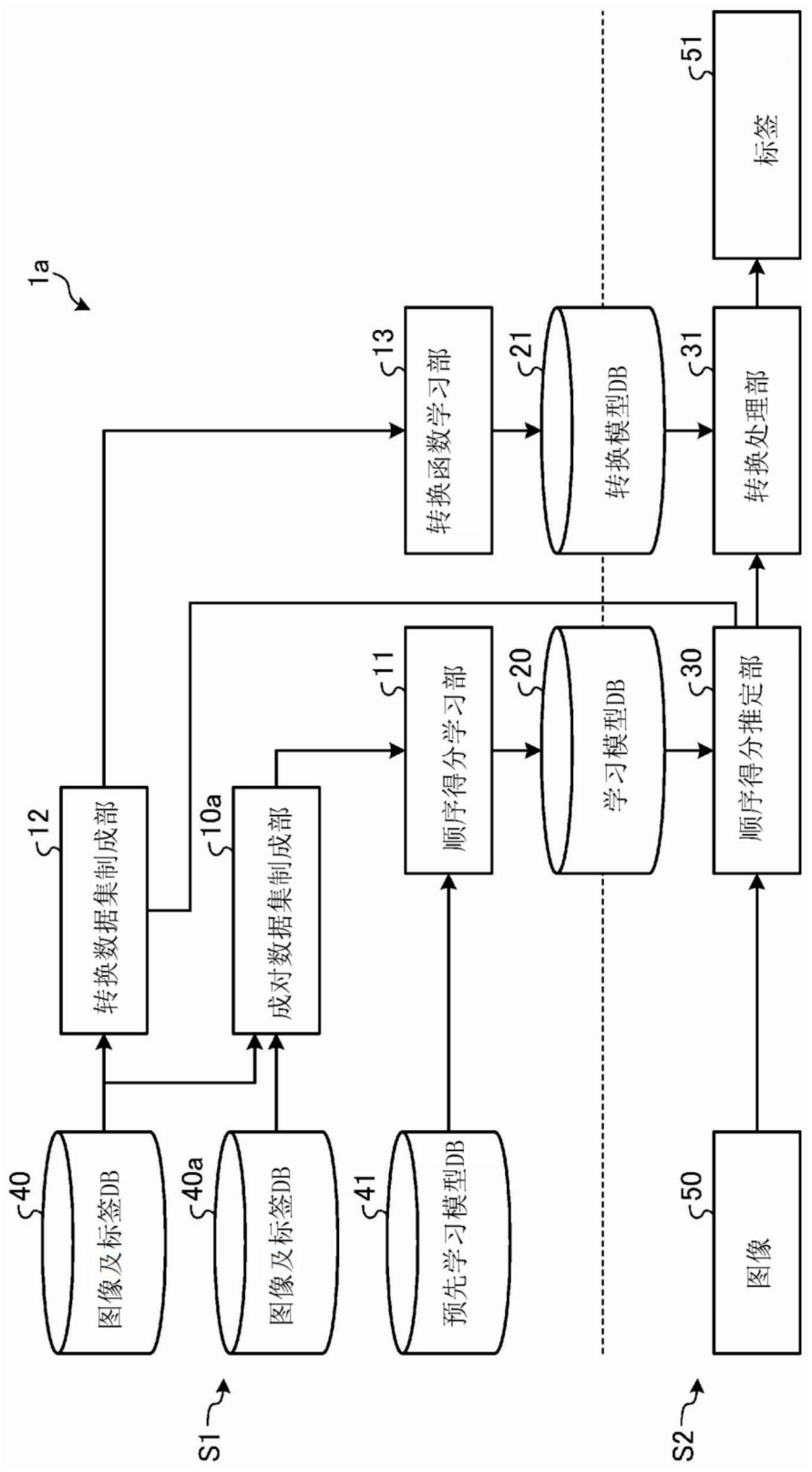

1a
40
图像及标签DB
40a
图像及标签DB
41
预先学习模型DB
12
转换数据集制成部
10a
成对数据集制成部
11
顺序得分学习部
13
转换函数学习部
20
学习模型DB
21
转换模型DB
30
顺序得分推定部
31
转换处理部
50
图像
51
标签
S1
S2

图5

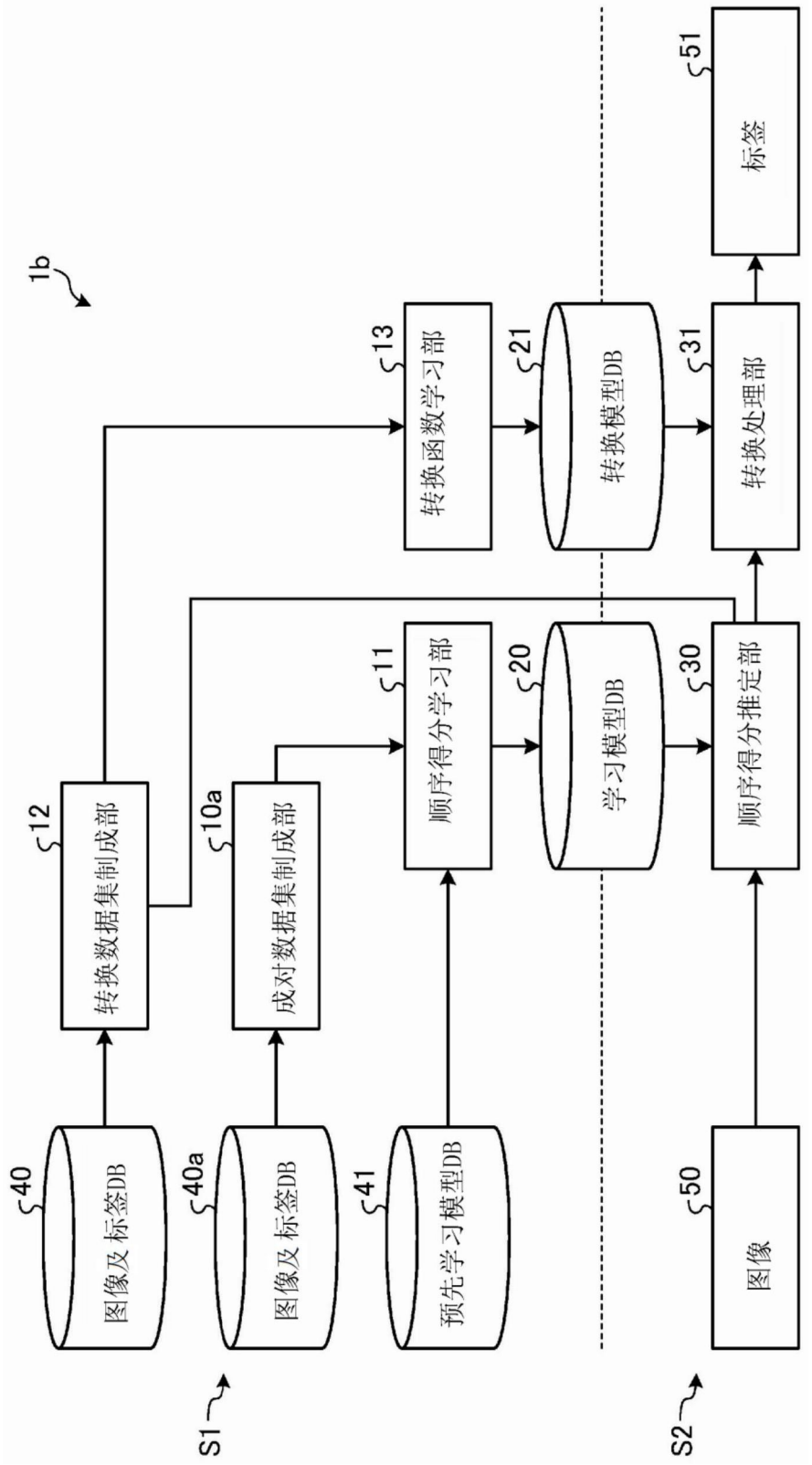

1b
40
图像及标签DB
40a
图像及标签DB
41
预先学习模型DB
S1
12
转换数据集制成部
10a
成对数据集制成部
11
顺序得分学习部
13
转换函数学习部
20
学习模型DB
21
转换模型DB
30
顺序得分推定部
31
转换处理部
50
图像
S2
51
标签

图6

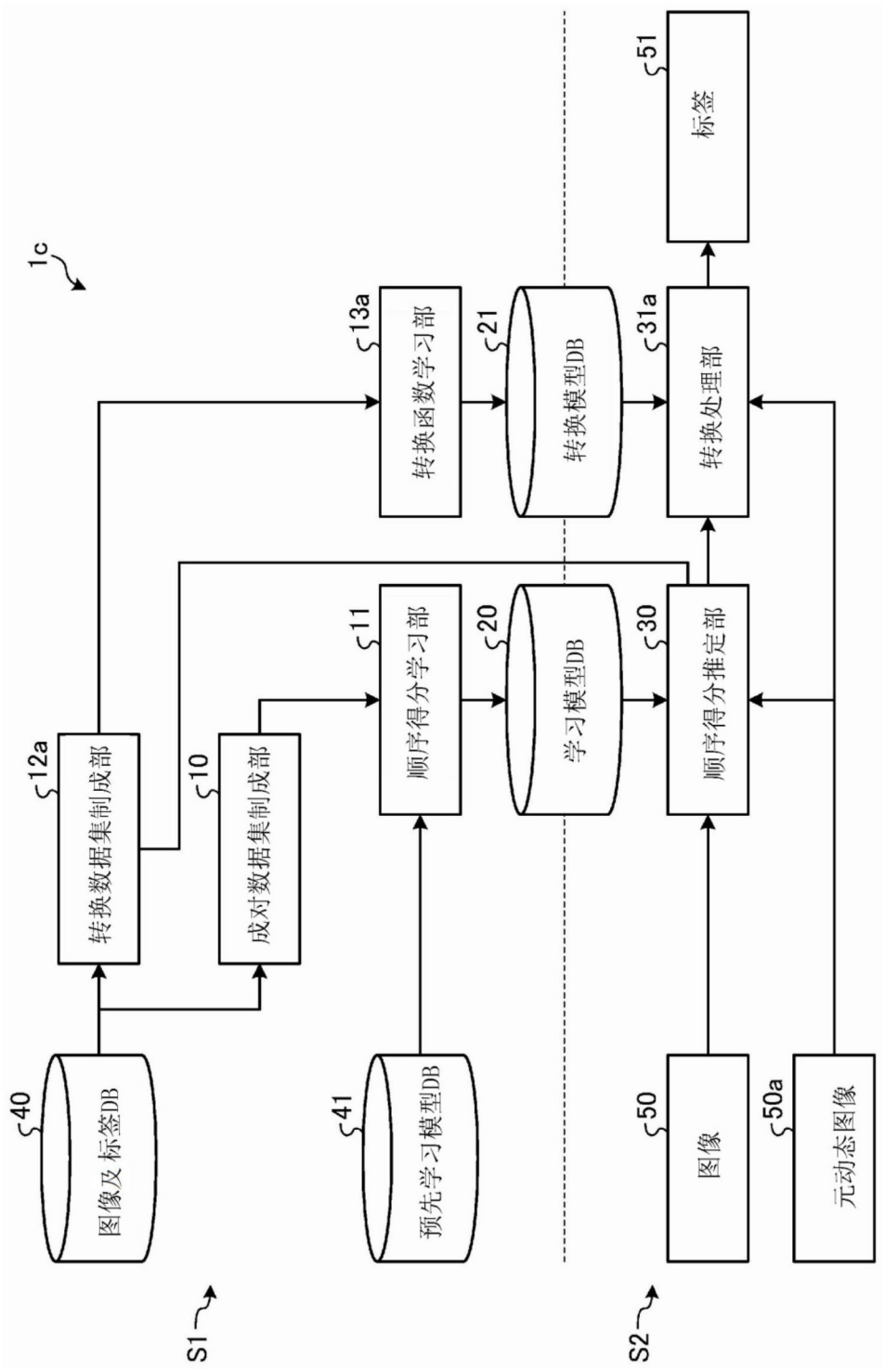

1c
13a
转换函数学习部
21
转换模型DB
31a
转换处理部
51
标签
12a
转换数据集制成部
10
成对数据集制成部
11
顺序得分学习部
20
学习模型DB
30
顺序得分推定部
40
图像及 标签DB
41
预先学习模型DB
50
图像
50a
元动态图像
S1
S2

图7

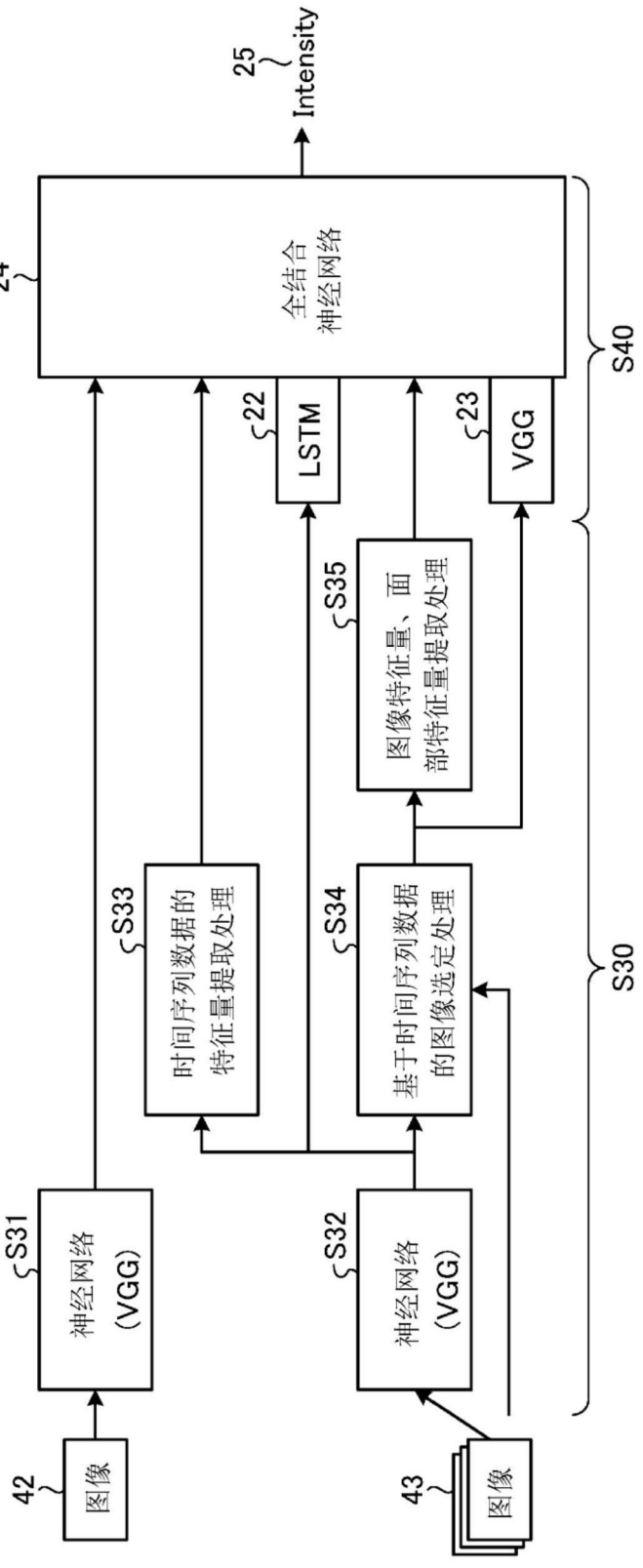

25
Intensity
24
全结合
神经网络
S40
22
LSTM
23
VGG
S35
图像特征量、面
部特征量提取处理
S33
时间序列数据的
特征量提取处理
S34
基于时间序列数据
的图像选定处理
S30
S31
神经网络
(VGG)
S32
神经网络
(VGG)
42
图像
43
图像

图8

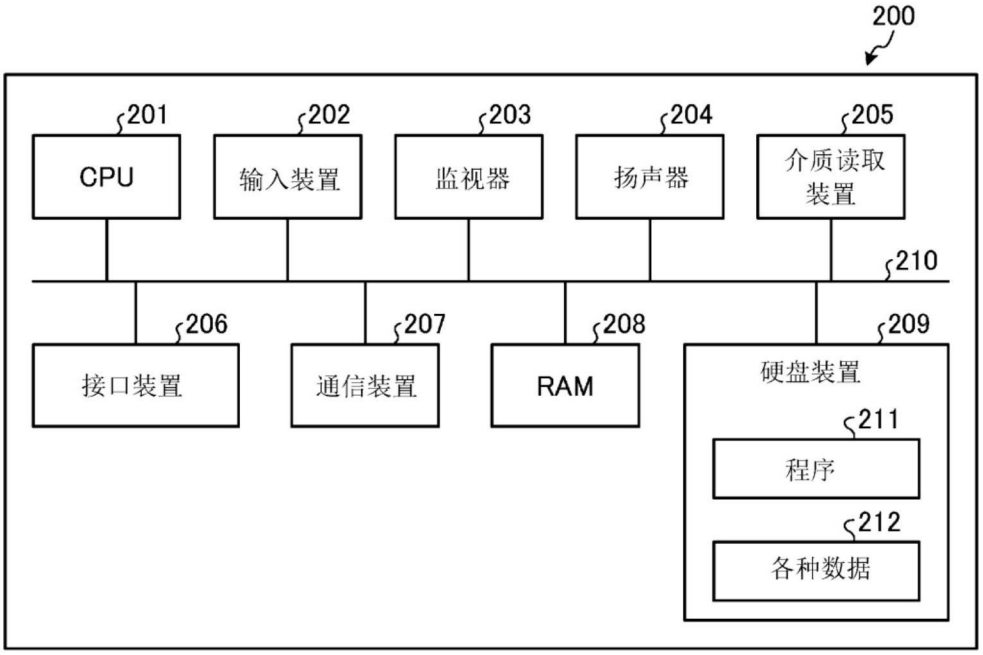

200
201
CPU
202
输入装置
203
监视器
204
扬声器
205
介质读取装置
210
206
接口装置
207
通信装置
208
RAM
209
硬盘装置
211
程序
212
各种数据

图9