(19)国家知识产权局

(12)发明专利

(10)授权公告号 CN 114298997 B
(45)授权公告日 2023.06.02

(21)申请号 202111593609.0

(22)申请日 2021.12.23

(65)同一申请的已公布的文献号
申请公布号 CN 114298997 A

(43)申请公布日 2022.04.08

(73)专利权人 北京瑞莱智慧科技有限公司
地址 100084 北京市海淀区中关村东路1号院8号楼19层A1901

(72)发明人 田天 请求不公布姓名
请求不公布姓名

(74)专利代理机构 北京箴思知识产权代理有限公司 11913
专利代理师 李春晖 曾晓波

(51)Int.Cl.
G06T 7/00(2017.01)
G06N 3/0464(2023.01)
G06N 3/0455(2023.01)
G06N 3/08(2023.01)

审查员 钟福煌

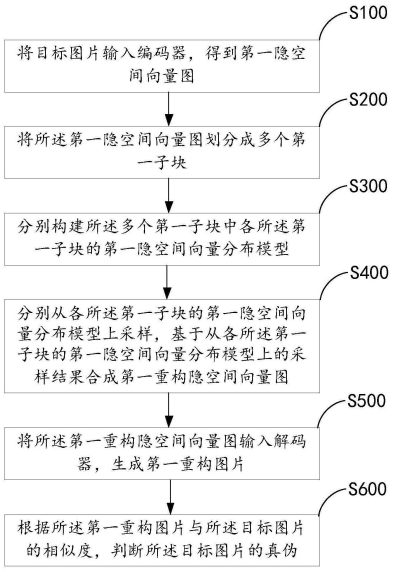
权利要求书4页 说明书19页 附图9页

(54)发明名称

一种伪造图片检测方法、装置及存储介质

(57)摘要

本申请涉及深度学习领域，公开一种伪造图片检测方法、装置及存储介质。该方法包括：将目标图片输入编码器，得到目标图片的第一隐空间向量图；将第一隐空间向量图划分成多个第一子块；分别构建多个第一子块中各第一子块的第一隐空间向量分布模型；分别从各第一子块的第一隐空间向量分布模型上采样，基于从各第一子块的第一隐空间向量分布模型上的采样结果合成第一重构隐空间向量图；将第一重构隐空间向量图输入解码器，生成第一重构图片；根据第一重构图片与目标图片的相似度，判断目标图片的真伪。该方法，在训练时无需获取伪造图片，在检测时能够在保留目标图片空间结构信息的前提下进行重构，因此在检测伪造图片时更加的精准。



CN 114298997 B

1.一种伪造图片检测方法,所述方法包括:

将目标图片输入编码器,得到所述目标图片的第一隐空间向量图,所述目标图片根据待检测图片或历史重构图片得到,所述历史重构图片根据历史重构隐空间向量得到;

将所述第一隐空间向量图划分成多个第一子块;

分别构建所述多个第一子块中各所述第一子块的第一隐空间向量分布模型;

分别从各所述第一子块的第一隐空间向量分布模型上采样,基于从各所述第一子块的第一隐空间向量分布模型上的采样结果合成第一重构隐空间向量图;

将所述第一重构隐空间向量图输入解码器,生成第一重构图片;

根据所述第一重构图片与所述目标图片的相似度,判断所述目标图片的真伪。

2.如权利要求1所述的伪造图片检测方法,所述分别构建所述多个第一子块中各所述第一子块的第一隐空间向量分布模型,包括:

基于所述第一隐空间向量图,分别计算各所述多个第一子块的隐空间向量的平均值和方差;

基于各所述多个第一子块的隐空间向量的平均值和方差,分别构建各所述第一子块的第一隐空间向量分布模型。

3.如权利要求2所述的伪造图片检测方法,所述分别从各所述第一子块的第一隐空间向量分布模型上采样,基于从各所述第一子块的第一隐空间向量分布模型上的采样结果合成第一重构隐空间向量图,包括:

分别从各所述第一子块的第一隐空间向量分布模型上采样,得到各所述第一子块的第一隐空间向量分布模型上的采样结果;

基于各所述第一子块的第一隐空间向量分布模型上的采样结果,合成与所述目标图片具有相同位置分布的所述第一重构隐空间向量图。

4.如权利要求3所述的伪造图片检测方法,所述根据所述第一重构图片与所述目标图片的相似度,判断所述目标图片的真伪,包括:

将所述第一重构图片输入至所述编码器,得到所述第一重构图片的第二隐空间向量图;

基于所述第一隐空间向量图和所述第二隐空间向量图的相似度,判断所述目标图片的真伪。

5.如权利要求4所述的伪造图片检测方法,所述基于所述第一隐空间向量图和所述第二隐空间向量图的相似度,判断所述目标图片的真伪,包括:

将所述第二隐空间向量图划分成与所述第一隐空间向量图相同的多个第二子块;

基于各个所述第二子块与各个所述第一子块的相似度,判断所述目标图片的真伪。

6.如权利要求5所述的伪造图片检测方法,所述基于各个所述第二子块与各个所述第一子块的相似度,判断所述目标图片的真伪,包括:

分别计算各个所述第二子块以及各个所述第一子块隐空间向量的方差和平均值;

基于各个所述第二子块以及各个所述第一子块隐空间向量的方差和平均值的相似度,判断所述待检测图片的真伪。

7.如权利要求5所述的伪造图片检测方法,所述基于各个所述第二子块与各个所述第一子块的相似度,判断所述目标图片的真伪,包括:

基于所述第二隐空间向量图，构建各个所述第二子块的第二隐空间向量分布模型；

从各所述第二子块的第二隐空间向量分布模型上采样，基于在各所述第二子块的第二隐空间向量分布模型上的采样结果，与在各所述第一子块的第一隐空间向量分布模型上的采样结果的相似度，判断所述目标图片的真伪。

8.如权利要求7所述的伪造图片检测方法，所述编码器和所述解码器经过如下训练得到：

获取真实的训练图片；

将所述训练图片输入所述编码器，得到所述训练图片的第三隐空间向量图；

将所述第三隐空间向量图划分成多个第三子块；

分别构建各所述第三子块的第三隐空间向量分布模型；

从各所述第三子块的第三隐空间向量分布模型上采样，合成第三重构隐空间向量图；

将所述第三重构隐空间向量图输入所述解码器，生成与所述训练图片对应的第三重构图片；

计算所述第三重构图片与对应的训练图片之间的像素重构损失和隐空间分布损失，基于所述像素重构损失和所述隐空间分布损失中，调整所述解码器的模型参数和所述编码器的模型参数，直至满足预设结束条件时结束训练。

9.如权利要求8所述的伪造图片检测方法，所述基于所述像素重构损失和所述隐空间分布损失中的至少一项，调整所述解码器的模型参数和所述编码器的模型参数，直至满足预设结束条件时结束训练，包括：

基于各所述第三子块的第三隐空间向量分布模型和标准高斯分布计算所述隐空间分布损失；

基于所述第三重构图片与对应的训练图片计算所述像素重构损失；

基于所述隐空间分布损失和所述像素重构损失，利用随机梯度下降算法对所述解码器的模型参数和所述编码器的模型参数进行更新，直至所述编码器的模型参数和所述解码器的模型参数收敛。

10.如权利要求1所述的伪造图片检测方法，所述根据所述第一重构图片与所述目标图片的相似度，判断所述目标图片的真伪，包括：

通过比较所述第一重构图片与所述目标图片的像素值差异，判断所述目标图片的真伪。

11.一种伪造图片检测装置，包括：

输入输出模块，用于向编码器输入目标图片；

编码器，用于对自所述输入输出模块输入的所述目标图片进行编码处理，得到所述目标图片的第一隐空间向量图，将所述第一隐空间向量图划分成多个第一子块，构建各所述第一子块的第一隐空间向量分布模型；

处理模块，用于从编码器构建的各所述第一子块的第一隐空间向量分布模型上采样，合成第一重构隐空间向量图；

解码器，用于将所述处理模块合成的第一重构隐空间向量图解码，生成与所述目标图片对应的第一重构图片；

所述处理模块，还用于通过比较所述解码器解码生成的所述第一重构图片与所述目标

图片的相似度，判断所述目标图片的真伪。

12.如权利要求11所述的伪造图片检测装置，所述编码器还用于，基于长、宽、通道中的至少一个维度将所述第一隐空间向量图划分成所述多个第一子块。

13.如权利要求11所述的伪造图片检测装置，所述编码器还用于，基于从所述目标图片提取的所述第一隐空间向量图，分别计算各所述第一子块的隐空间向量的平均值和方差；以及

基于各所述第一子块的隐空间向量的平均值和方差，分别构建各所述第一子块的第一隐空间向量分布模型。

14.如权利要求13所述的伪造图片检测装置，所述处理模块还用于，从所述编码器构建的各所述第一子块的第一隐空间向量分布模型上采样，并合成与所述目标图片具有相同位置分布的所述第一重构隐空间向量图。

15.如权利要求11所述的伪造图片检测装置，所述处理模块还用于，将所述解码器生成的第一重构图片输入至所述编码器；

所述编码器还用于，对所述第一重构图片进行提取第二隐空间向量图；

所述处理模块还用于，通过比较所述第一隐空间向量图和所述第二隐空间向量图的相似度，判断所述目标图片的真伪。

16.如权利要求15所述的伪造图片检测装置，所述解码器还用于，将所述第二隐空间向量图划分成与所述第一隐空间向量图相同的多个第二子块；

所述处理模块还用于，通过比较各个所述第二子块与各个所述第一子块的相似度，判断所述目标图片的真伪。

17.如权利要求16所述的伪造图片检测装置，所述编码器还用于，分别计算各个所述第二子块以及各个所述第一子块隐空间向量的方差和平均值；

所述处理模块还用于，基于各个所述第二子块以及各个所述第一子块隐空间向量的方差和平均值的相似度，判断所述目标图片的真伪。

18.如权利要求16所述的伪造图片检测装置，所述编码器还用于，基于所述第二隐空间向量图，构建各个所述第二子块的第二隐空间向量分布模型；

所述处理模块还用于，从各所述第二子块的第二隐空间向量分布模型上采样；以及

基于从各所述第二子块的第二隐空间向量分布模型上的采样结果，与在各所述第一子块的第一隐空间向量分布模型上的采样结果的相似度，判断所述目标图片的真伪。

19.如权利要求11-18任一项所述的伪造图片检测装置，所述处理模块还用于，通过比较所述解码器生成的所述第一重构图片与所述目标图片的像素值相似度，判断所述目标图片的真伪。

20.如权利要求11-18任一项所述的伪造图片检测装置，所述解码器和所述编码器经过如下训练得到：

获取真实的训练图片；

通过所述输入输出模块将所述训练图片输入所述编码器，得到所述训练图片的第三隐空间向量图；

通过编码器将所述第三隐空间向量图划分成多个第三子块；

通过编码器分别构建各所述第三子块的第三隐空间向量分布模型；

通过处理模块从各所述第三子块的第三隐空间向量分布模型上采样，合成第三重构隐空间向量图；

通过处理模块将所述第三重构隐空间向量图输入所述解码器，通过所述解码器对所述第三重构进空间向量图进行解码，生成与所述训练图片对应的第三重构图片；

通过所述处理模块计算所述第三重构图片与对应的训练图片之间的像素重构损失和隐空间分布损失，基于所述像素重构损失和所述隐空间分布损失，调整所述编码器的模型参数和所述解码器的模型参数，直至满足预设结束条件时结束训练。

21.如权利要求20所述的伪造图片检测装置，所述处理模块还用于：

基于所述编码器构建的各所述第三子块的第三隐空间向量分布模型和标准高斯分布计算所述隐空间分布损失；

基于所述解码器解码生成的第三重构图片与对应的训练图片计算所述像素重构损失；以及

基于所述隐空间分布损失和所述像素重构损失，利用随机梯度下降算法对所述编码器的模型参数和所述解码器的模型参数进行更新，直至所述解码器的模型参数和所述编码器的模型参数收敛。

22.一种计算机可读存储介质，其包括指令，当其在计算机上运行时，使得计算机执行如权利要求1-10中任一项所述的方法。

23.一种处理设备，所述处理设备包括：

至少一个处理器、存储器和输入输出单元；

其中，所述存储器用于存储计算机程序，所述处理器用于调用所述存储器中存储的计算机程序来执行如权利要求1-10中任一项所述的方法。

一种伪造图片检测方法、装置及存储介质

技术领域

[0001] 本申请实施例涉及深度学习领域，特别涉及一种伪造图片检测方法、装置及存储介质。

背景技术

[0002] 目前，在深度伪造检测领域，对于伪造图片的检测，一种是通过收集大量的真实图片和伪造图片，然后基于真实图片和伪造图片对二分类深度神经网络进行训练，并使用训练后的二分类网络对伪造图片进行检测。此种方法需要获取大量的真实图片和伪造图片，真实图片的获取可以基于互联网或者一些开源的数据源集，相对比较容易获得，但对于伪造图片而言如要获取大量的数据则相对比较困难。

[0003] 虽然可以利用变分自动编码器进行检测，且变分自动编码器可以仅基于真实图片就能完成训练过程，比如常见的变分自动编码器通过提取真实图片的隐空间向量，并将其摊平转换成一维度向量，然后基于摊平的一维度向量构造真实图片的隐空间向量分布模型，基于隐空间向量分布模型生成重构图片。

[0004] 但是，该变分自动编码器在将真实图片的隐空间向量摊平转换成一维度向量时，会破坏真实图片本身的空间结构信息，从而导致基于摊平后的一维度向量构造的隐空间向量分布模型并不能真实的反映真实图片的空间结构信息，故而基于该隐空间向量分布模型重构的图片也不准确，最终导致训练后的变分自动编码器在检测伪造图片时也具有较大的检测隐患。

发明内容

[0005] 本申请实施例提供一种伪造图片检测方法、装置及存储介质，能够在保证不丧失真实图片空间结构信息的前提下，对变分自动编码器进行训练，从而利用训练后的变分自动编码器生成的重构图片也能保持与真实图片相同的空间结构信息，以提高检测伪造图片的准确率。

[0006] 第一方面，本申请实施例提出一种伪造图片检测方法，所述方法应用于深度学习中的伪造图片检测模型，或者伪造图片生成模型，所述方法包括：

[0007] 将目标图片输入编码器，得到所述目标图片的第一隐空间向量图，所述目标图片根据待检测图片或历史重构图片得到，所述历史重构图片根据历史重构隐空间向量得到；

[0008] 将所述第一隐空间向量图划分成多个第一子块；

[0009] 分别构建所述多个第一子块中各所述第一子块的第一隐空间向量分布模型；

[0010] 分别从各所述第一子块的第一隐空间向量分布模型上采样，基于从各所述第一子块的第一隐空间向量分布模型上的采样结果合成第一重构隐空间向量图；

[0011] 将所述第一重构隐空间向量图输入解码器，生成第一重构图片；

[0012] 根据所述第一重构图片与所述目标图片的相似度，判断所述目标图片的真伪。

[0013] 一种可能的设计中，所述多个第一子块基于长、宽、通道中的至少一个维度划分。

[0014] 第二方面，本申请实施例提出一种伪造图片检测装置，具有实现对应于上述第一方面提供的伪造图片检测方法的功能。所述功能可以通过硬件实现，也可以通过硬件执行相应的软件实现。硬件或软件包括一个或多个与上述功能相对应的模块，所述模块可以是软件和/或硬件。

[0015] 一种可能的设计中，所述伪造图片检测装置包括：

[0016] 输入输出模块，用于向编码器输入目标图片；

[0017] 编码器，用于对自所述输入输出模块输入的所述目标图片进行编码处理，得到所述目标图片的第一隐空间向量图，将所述第一隐空间向量图划分成多个第一子块，构建各所述第一子块的第一隐空间向量分布模型；

[0018] 处理模块，用于从所述编码器构建的各所述第一子块的第一隐空间向量分布模型上采样，合成第一重构隐空间向量图；

[0019] 解码器，用于将所述处理模块合成的第一重构隐空间向量图解码生成与所述目标图片对应的第一重构图片；

[0020] 所述处理模块，还用于通过比较所述解码器解码生成的所述第一重构图片与所述目标图片的相似度，判断所述目标图片的真伪。

[0021] 一种可能的设计中，所述编码器还用于：基于长、宽、通道中的至少一个维度将所述第一隐空间向量图划分成所述多个第一子块。

[0022] 在本申请的另一个实施例中，所述编码器用于：基于从所述目标图片提取的所述第一隐空间向量图，分别计算各所述多个第一子块的隐空间向量的平均值和方差；

[0023] 基于各所述多个第一子块的隐空间向量的平均值和方差，分别构建各所述第一子块的第一隐空间向量分布模型。

[0024] 一种可能的设计中，所述处理模块用于从所述编码器构建的各所述第一子块的第一隐空间向量分布模型上采样，合成与所述目标图片具有相同位置分布的所述第一重构隐空间向量图。

[0025] 一种可能的设计中，所述处理模块用于：

[0026] 将所述解码器生成的第一重构图片输入至所述编码器，所述编码器对所述第一重构图片提取第二隐空间向量图；

[0027] 所述处理模块通过比较所述第一隐空间向量图和所述第二隐空间向量图的相似度，判断所述目标图片的真伪。

[0028] 一种可能的设计中，所述编码器将所述第二隐空间向量图划分成与所述第一隐空间向量图相同的多个第二子块，所述处理模块通过比较各个所述第二子块与各个所述第一子块的相似度，判断所述目标图片的真伪。

[0029] 一种可能的设计中，所述编码器分别计算各个所述第二子块以及各个所述第一子块隐空间向量的方差和平均值，所述处理模块通过比较各个所述第二子块以及各个所述第一子块隐空间向量的方差和平均值的相似度，判断所述目标图片的真伪。

[0030] 一种可能的设计中，所述编码器基于所述第二隐空间向量图，构建各个所述第二子块的第二隐空间向量分布模型；

[0031] 所述处理模块从各所述第二子块的第二隐空间向量分布模型上采样；

[0032] 所述处理模块通过比较从各所述第二子块的第二隐空间向量分布模型上的采样

结果，与在各所述第一子块的第一隐空间向量分布模型上的采样结果的相似度，判断所述目标图片的真伪。

[0033] 一种可能的设计中，所述处理模块通过比较所述解码器生成的所述第一重构图片与所述目标图片的像素值相似度，判断所述目标图片的真伪。

[0034] 一种可能的设计中，所述解码器和所述编码器经过如下训练得到：

[0035] 获取真实的训练图片；

[0036] 通过所述输入输出模块将所述训练图片输入所述编码器200，得到所述训练图片的第三隐空间向量图；

[0037] 通过编码器将所述第三隐空间向量图划分成多个第三子块；

[0038] 通过编码器分别构建各所述第三子块的第三隐空间向量分布模型；

[0039] 处理模块从各所述第三子块的第三隐空间向量分布模型上采样，合成第三重构隐空间向量图；

[0040] 处理模块将所述第三重构隐空间向量图输入所述解码器，所述解码器对所述第三重构进空间向量图进行解码，生成与所述训练图片对应的第三重构图片；

[0041] 处理模块计算所述第三重构图片与对应的训练图片之间的像素重构损失和隐空间分布损失，基于所述像素重构损失和所述隐空间分布损失，调整所述编码器的模型参数和所述解码器的模型参数，直至满足预设结束条件时结束训练。

[0042] 一种可能的设计中，所述处理模块还用于：

[0043] 基于所述编码器构建的各所述第三子块的第三隐空间向量分布模型和标准高斯分布计算所述隐空间分布损失；

[0044] 基于所述解码器解码生成的第三重构图片与对应的训练图片计算所述像素损失；

[0045] 基于所述隐空间分布损失和所述像素损失，利用随机梯度下降算法对所述编码器的模型参数和所述解码器的模型参数进行更新，直至所述解码器的模型参数和所述编码器的模型参数收敛。

[0046] 第三方面，本申请实施例提出一种计算机可读存储介质，其包括指令，当其在计算机上运行时，使得计算机执行上述第一方面、或者第一方面的可能的设计中所述的方法。

[0047] 第四方面，本申请实施例提出一种处理设备，所述处理设备包括：

[0048] 至少一个处理器、存储器和输入输出单元；

[0049] 其中，所述存储器用于存储计算机程序，所述处理器用于调用所述存储器中存储的计算机程序来执行上述第一方面、或者第一方面的可能的设计中所述的方法。

[0050] 相较于现有技术中的变分自动编码器对目标图片的隐空间向量图进行摊平操作得到一维度向量图，然后再基于摊平后的一维度向量图进行构建隐空间向量分布模型，在摊平操作时破坏了目标图片自身的空间结构信息，故而构建的隐空间向量分布模型并不准确，那么基于此隐空间向量分布模型进行取样重构的图片也不具有目标图片的空间结构信息，基于此与目标图片进行对比判断真伪的准确率较低。而本申请实施例中，是通过将目标图片的隐空间向量图划分成多个第一子块，然后再基于划分后的多个第一子块分别构建隐空间向量分布模型，在划分多个第一子块时，各个第一子块能够保留目标图片自身的空间结构信息，故而基于各个第一子块构建的各个隐空间向量分布模型也能够保留目标图片自身的空间结构信息，基于从各个隐空间向量分布模型的采样结果，第一重构隐空间向量分

布图也能够保留目标图片的空间结构信息，那么基于第一重构隐空间向量图解码生成的第一重构图片也与目标图片具有相同的空间结构信息，基于第一重构图片与目标图片进行对比相似度，判断目标图片真伪的准确率较高。

附图说明

[0051] 为了更清楚地说明本申请实施例或现有技术中的技术方案，下面将对实施例或现有技术描述中所需要使用的附图作简单地介绍，显而易见地，下面描述中的附图仅仅是本申请的一些实施例，对于本领域普通技术人员来讲，在不付出创造性劳动的前提下，还可以根据这些附图示出的结构获得其他的附图。

[0052] 图1为当前伪造图片检测装置的检测流程示意图；

[0053] 图2a为本申请实施例中实施伪造图片检测方法的神经网络模型的一种内部处理示意图；

[0054] 图2b为本申请实施例中实施伪造图片检测方法所涉及的一种伪造图片检测模型的结构示意图；

[0055] 图3为本申请实施例中伪造图片检测方法的一种流程示意图；

[0056] 图4a为本申请实施例伪造图片检测方法在一伪造图片检测场景的应用图；

[0057] 图4b为本申请实施例伪造图片检测方法在另一伪造图片检测场景的应用图；

[0058] 图5为本申请实施例伪造图片检测方法中对于第一向量图一实施例的划分示意图；

[0059] 图6为本申请实施例伪造图片检测方法中对于第一向量图另一实施例的划分示意图；

[0060] 图7为本申请实施例伪造图片检测方法中对于第一向量图另一实施例的划分示意图；

[0061] 图8为本申请实施例伪造图片检测方法中对于第一向量图另一实施例的划分示意图；

[0062] 图9为本申请实施例伪造图片检测方法中对于第一向量图另一实施例的划分示意图；

[0063] 图10为本申请实施例伪造图片检测方法中对于第一向量图另一实施例的划分示意图；

[0064] 图11为本申请实施例伪造图片检测方法中对于第一向量图另一实施例的划分示意图；

[0065] 图12为利用本申请实施例伪造图片检测方法，对真实人脸图片和伪造人脸图片进行重构后的对比图；

[0066] 图13本申请实施例中伪造图片检测装置的一种结构示意图；

[0067] 图14为本申请一实施例中实施伪造图片检测方法的实体装置的结构图；

[0068] 图15为本申请一实施例中实施伪造图片检测方法的服务器的结构示意图。

具体实施方式

[0069] 下面将参考若干示例性实施方式来描述本申请实施例的原理和精神。应当理解，

给出这些实施方式仅仅是为了使本领域技术人员能够更好地理解进而实现本申请实施例，而并非以任何方式限制本申请实施例的范围。

[0070] 本领域技术人员知道，本申请实施例的实施方式可以实现为一种装置、设备、方法或计算机程序产品。因此，本公开可以具体实现为以下形式，即：完全的硬件、完全的软件(包括固件、驻留软件、微代码等)，或者硬件和软件结合的形式。

[0071] 需要说明的是，本发明的说明书和权利要求书及上述附图中的术语“第一”、“第二”等是用于区别类似的对象，而不必用于描述特定的顺序或先后次序。应该理解这样使用的数据在适当情况下可以互换，以便这里描述的本发明的实施例能够以除了在这里图示或描述的那些以外的顺序实施。此外，术语“包括”和“具有”以及他们的任何变形，意图在于覆盖不排他的包含，例如，包含了一系列步骤或单元的过程、方法、系统、产品或设备不必限于清楚地列出的那些步骤或单元，而是可包括没有清楚地列出的或对于这些过程、方法、产品或设备固有的其它步骤或单元。

[0072] 本申请实施例提供一种伪造图片检测方法、装置及存储介质，可以运用于图片检测领域、人脸识别领域以及图片识别模型和人脸识别模型的训练领域等场景。该方案可用于伪造图片检测装置，该伪造图片检测装置可部署于服务器侧或终端侧，本申请实施例不对此作限定，后续以伪造图片检测装置部署于服务器侧为例实施该伪造图片检测方法。

[0073] 本申请实施例提供的方案涉及人工智能(Artificial Intelligence，AI)、自然语言处理(Nature Language processing，NLP)、机器学习(Machine Learning，ML)等技术，具体通过如下实施例进行说明：

[0074] 其中，人工智能(Artificial Intelligence，AI)是利用数字计算机或者数字计算机控制的机器模拟、延伸和扩展人的智能，感知环境、获取知识并使用知识获得最佳结果的理论、方法、技术及应用系统。换句话说，人工智能是计算机科学的一个综合技术，它企图了解智能的实质，并生产出一种新的能以人类智能相似的方式做出反映的智能机器。人工智能也就是研究各种智能机器的设计原理与实现方法，使机器具有感知、推理与决策的功能。

[0075] 人工智能技术是一门综合学科，涉及领域广泛，既有硬件层面的技术也有软件层面的技术。人工智能基础技术一般包括如传感器、专用人工智能芯片、云计算、分布式存储、大数据处理技术、操作/交互系统、机电一体化等技术。人工智能软件技术主要包括计算机视觉技术、语音处理技术、自然语言处理技术以及机器学习/深度学习等几大方向。

[0076] 计算机视觉技术(Computer Vision，CV)计算机视觉是一门研究如何使机器“看”的科学，更进一步的说，就是指用摄影机和电脑代替人眼对目标进行识别、跟踪和测量等机器视觉，并进一步做图形处理，使电脑处理成为更适合人眼观察或传送给仪器检测的图像。作为一个科学学科，计算机视觉研究相关的理论和技术，试图建立能够从图像或者多维数据中获取信息的人工智能系统。计算机视觉技术通常包括图像处理、图像识别、图像语义理解、图像检索、OCR、视频处理、视频语义理解、视频内容/行为识别、三维物体重建、3D技术、虚拟现实、增强现实、同步定位与地图构建等技术，还包括常见的人脸识别、指纹识别等生物特征识别技术。

[0077] 机器学习(Machine Learning，ML)是一门多领域交叉学科，涉及概率论、统计学、逼近论、凸分析、算法复杂度理论等多门学科。专门研究计算机怎样模拟或实现人类的学习行为，以获取新的知识或技能，重新组织已有的知识结构使之不断改善自身的性能。机器学

习是人工智能的核心，是使计算机具有智能的根本途径，其应用遍及人工智能的各个领域。机器学习和深度学习通常包括人工神经网络、置信网络、强化学习、迁移学习、归纳学习、式教学习等技术。

[0078] 随着人工智能技术研究和进步，人工智能技术在多个领域展开研究和应用，例如常见的智能家居、智能穿戴设备、虚拟助理、智能音箱、智能营销、无人驾驶、自动驾驶、无人机、机器人、智能医疗、智能客服等，相信随着技术的发展，人工智能技术将在更多的领域得到应用，并发挥越来越重要的价值。

[0079] 一些实施方式中，如图2a所示，图2a为本申请实施例中实施伪造图片检测方法所涉及的一种通信系统框架示意图。该通信系统可以包括至少一个终端和至少一个服务器，本申请实施例以一个终端01和一个服务器02为例。

[0080] 该服务器02中部署伪造图片检测模型，如图2b所示，为伪造图片检测模型的结构示意，包括输入输出模块、编码器、处理模块以及解码器。

[0081] 输入输出模块用于供用户输入目标图片。

[0082] 编码器，可以由一系列的卷积神经网络组成，通过卷积神经网络可以针对所输入的目标图片，进行提取第一隐空间向量图，对于一张图片而言，图片在每个位置所展现出的颜色、尺寸、图形等特征均由该位置的隐空间向量所决定，各个位置的隐空间向量组成了该图片的隐空间向量图。编码器可以部署在计算机、笔记本、手机、平板、扫描仪等设备上，比如将编码器设置的计算机上，计算机通过预设的接口获取目标图片后，通过网络将目标图片发送至编码器，进而编码器对目标图片进行提取第一隐空间向量图。编码器还可以对第一隐空间向量图进行划分成多个第一子块，构建各个第一子块的第一隐空间向量分布模型。

[0083] 处理模块，可以在编码器所构建的各个第一隐空间向量分布模型上采样，并基于各个采样结果合成，第一重构隐空间向量图。

[0084] 解码器可以对采样器合成的第一重构隐空间向量图进行解码，生成第一重构图片。

[0085] 另外，处理模块还可以基于解码器生成的第一重构图片以及目标图片的相似度，对目标图片的真伪进行判断。

[0086] 用户有伪造图片检测需求时，可通过终端01向该服务器02发送待检测图片，服务器02可以采用伪造图片检测模型对该待检测图片进行重构。服务器02可以向终端01反馈重构图片，终端01进而可以基于目标图片以及该重构图片，进行判断该待检测图片的真伪；或者服务器02可以直接根据待检测图片和重构图片，判断待检测图片的真伪，然后直接将将判断结果发送至终端01。

[0087] 其中，需要特别说明的是，本申请实施例涉及的服务器可以是独立的物理服务器，也可以是多个物理服务器构成的服务器集群或者分布式系统，还可以是提供云计算服务的云服务器。终端可以是智能手机、平板电脑、笔记本电脑、台式计算机、智能音箱、智能手表等，但并不局限于此。终端以及服务器可以通过有线或无线通信方式进行直接或间接地连接，本申请在此不做限制。本申请实施例涉及的用户设备可以是智能手机、平板电脑、笔记本电脑、台式计算机、智能音箱、智能手表等，但并不局限于此。用户设备以及服务器可以通过有线或无线通信方式进行直接或间接地连接，本申请实施例在此不做限制。

[0088] 发明人研究发现，在伪造图片检测领域，对于伪造图片的检测，一种是通过收集大量的真实图片和伪造图片，然后基于真实图片和伪造数据对二分类深度神经网络进行训练，并使用训练后的二分类网络对伪造图片进行检测。此种方法需要获取大量的真实图片和伪造图片，真实图片的获取可以基于互联网或者一些开源的数据源集，相对比较容易获得，但对于伪造图片而言如要获取大量的数据则相对比较困难。另外还有一种检测方法是利用变分自动编码器，如图1所示，常见的变分自动编码器可以仅基于真实图片就能完成训练过程，其包括了编码器、采样器和解码器，编码器，可以由一系列的卷积神经网络组成，通过卷积神经网络可以针对所输入的目标图片，进行提取第一隐空间向量图，对于一张图片而言，图片在每个位置所展现出的颜色、尺寸、图形等特征均由该位置的隐空间向量所决定，各个位置的隐空间向量组成了该图片的隐空间向量图。编码器可以部署在计算机、笔记本、手机、平板、扫描仪等设备上，比如将编码器设置的计算机上，计算机通过预设的接口获取目标图片后，通过网络将目标图片发送至编码器，进而编码器对目标图片进行提取第一隐空间向量图。向编码器输入目标图片，编码器提取目标图片的隐空间向量图，并将其摊平转换成一维度向量，然后基于摊平的一维度向量构造真实图片的隐空间向量分布模型，基于隐空间向量分布模型生成重构隐空间向量图，而后再通过解码器将重构隐空间向量图转换成重构图片并输出。对于此种常见的变分自动编码器而言，在将真实图片的隐空间向量摊平转换成一维度向量时，破坏了真实图片本身的空间结构信息，因此，基于摊平后的一维度向量构造的隐空间向量分布模型并不能反映真实图片的空间结构信息，故而重构图片也是不准确的，那么基于此种方法训练后的变分自动编码在检测伪造图片时具有较大的检测隐患。

[0089] 参考图3，以下介绍本申请实施例所提供的一种伪造图片检测方法，该方法由伪造图片检测装置执行，本申请实施例包括：

[0090] 步骤S100：将目标图片输入编码器，得到第一隐空间向量图。

[0091] 如图4a所示，目标图片可以为待检测图片，比如在本申请的一些实施例中，想要判断一张人脸图片是真实的人脸图片，还是伪造的人脸图片，则该人脸图片即为目标图片。另外目标图片还可以是历史重构图片(例如第二重构图片)得到。

[0092] 编码器可以针对所输入的目标图片，进行提取第一隐空间向量图，对于一张图片而言，图片在每个位置所展现出的颜色、尺寸、图形等特征均由该位置的隐空间向量所决定，各个位置的隐空间向量组成了该图片的隐空间向量图。

[0093] 步骤S200：将所述第一隐空间向量图划分成多个第一子块。

[0094] 第一隐空间向量图代表了目标图片各个位置的特征，在本步骤中，基于第一隐空间向量图，对其进行分割成多个第一子块，多个第一子块组合起来形成该第一隐空间向量图。分割时可以记录各个第一子块彼此之间的位置关系，根据各个第一子块的位置关系，能够将各个第一子块重新合成第一隐空间向量图。

[0095] 在本申请的一个实施例中，对于步骤S200将所述第一隐空间向量图划分成多个第一子块的步骤包括：基于长、宽、通道中的至少一个维度将所述第一隐空间向量图划分成所述多个第一子块。其中，长、宽为第一隐空间向量图的长度和宽度维度，通道为第一隐空间图所反映的目标图片像素颜色的维度，一般来说每个像素均有三个通道维度。

[0096] 接下来，用[C,H,W]代表目标图片输入编码器后提取到的第一隐空间向量图的通

道数，以及长宽尺寸，其中C为通道数，H为长度尺寸，W为宽度尺寸，假设通道数为3，长宽均为2，则第一隐空间向量图可以表示为[3,2,2]。以该第一隐空间向量图为例，对于基于长、宽、通道中的至少一个维度将所述第一隐空间向量图划分成所述多个第一子块进行阐述。

[0097] 1、基于单一长度维度，将所述第一隐空间向量图划分成多个第一子块。

[0098] 如图5所示，第一隐空间向量图[3,2,2]，那么按照单一维度长进行划分，则可以得到两个第一子块[3,1,2]。可知，得到的两个第一子块可以在长度维度上保持目标图片的空间结构信息。

[0099] 2、基于单一宽度维度，将所述第一隐空间向量图划分成多个第一子块。

[0100] 如图6所示，第一隐空间向量图[3,2,2]，那么按照单一维度宽进行划分，则可以得到两个第一子块[3,2,1]。可知，得到的两个第一子块可以在宽度维度上保持目标图片的空间结构信息。

[0101] 3、基于单一通道维度，将所述第一隐空间向量图划分成多个第一子块。

[0102] 如图7所示，第一隐空间向量图[3,2,2]，那么按照单一维度通道则可以得到三个第一子块[1,2,2]。可知，得到的三个第一子块可以在通道维度上保持目标图片的空间结构信息。

[0103] 4、基于长和宽两个维度，将所述第一隐空间向量图划分成多个第一子块。

[0104] 如图8所示，第一隐空间向量图[3,2,2]，那么按照维度长和宽进行划分，则可以得到四个第一子块[3,1,1]。可知，得到的四个第一子块可以在长度和宽度维度上保持目标图片的空间结构信息。

[0105] 5、基于长和通道两个维度，将所述第一隐空间向量图划分成多个第一子块。

[0106] 如图9所示，第一隐空间向量图[3,2,2]，那么按照维度长和通道进行划分，则可以得到六个第一子块[1,1,2]。可知，得到的六个第一子块可以在长度维度和通道维度上保持目标图片的空间结构信息。

[0107] 6、基于宽度和通道两个维度，将所述第一隐空间向量图划分成多个第一子块。

[0108] 如图10所示，第一隐空间向量图[3,2,2]，那么按照维度宽和通道进行划分，则可以得到六个第一子块[1,2,1]。可知，得到的六个个第一子块可以在宽度维度和通道维度上保持目标图片的空间结构信息。

[0109] 7、基于长度、宽度以及通道三个维度，将所述第一隐空间向量图划分成多个第一子块。

[0110] 如图11所示，第一隐空间向量图[3,2,2]，那么按照维度长、宽以及通道进行划分，则可以得到九个第一子块[1,1,1]。可知，得到的九个第一子块可以在长度、宽度以及通道三个维度上均保持目标图片的空间结构信息。

[0111] 步骤S300：分别构建所述多个第一子块中各所述第一子块的第一隐空间向量分布模型。

[0112] 隐空间向量分布模型可以为一个函数，该函数可以为高斯分布，每一个第一子块的第一隐空间向量分布模型即代表了该第一子块的隐空间向量所在的高斯分布，比如目标图片为人脸图片，对于基于“眼睛”位置的隐空间向量所构建的隐空间向量分布模型，就代表了人脸中的“眼睛”的隐空间向量所在的高斯分布，换言之，从“眼睛”位置的隐空间向量分布模型上取样，取样结果反馈在图片上也是“眼睛”，从“鼻子”位置的隐空间向量分布模

型上取样,取样结果反馈在图片上也是“鼻子”。

[0113] 在本申请的一个实施例中,对于步骤S300,分别构建所述多个第一子块中各所述第一子块的第一隐空间向量分布模型,包括如下步骤:

[0114] 步骤S310:基于所述第一隐空间向量图,分别计算各所述多个第一子块的隐空间向量的平均值和方差。

[0115] 如图4a所示,假设目标图片输入编码器后提取到的第一隐空间向量图为[3,2,2],并在步骤S200中按照长度和宽度两个维度对第一隐空间向量图进行划分,那么能够得到四个[3,1,1]的第一子块,在本步骤中就可以分别基于四个[3,1,1]的第一子块,分别进行计算各自的方差和平均值,在得到四个第一子块的方差和平均值之后,就可以利用每个第一子块的方差和平均值,分别构建各个第一子块的第一隐空间向量分布模型,如高斯分布,即如下步骤S320:基于各所述多个第一子块的隐空间向量的平均值和方差,分别构建各所述第一子块的第一隐空间向量分布模型。

[0116] 步骤S400:分别从各所述第一子块的第一隐空间向量分布模型上采样,基于从各所述第一子块的第一隐空间向量分布模型上的采样结果合成第一重构隐空间向量图。

[0117] 在步骤S300中具体阐述了每个第一子块的隐空间向量与各自的第一隐空间向量分布模型的关系,在步骤S400中,从各个第一子块的第一隐空间向量模型上采样,再按照各个第一子块之间的位置关系,重新组合起来,就能形成与第一隐空间向量图位置相同(各个第一子块排列位置相同)的第一重构隐空间向量图。

[0118] 在本申请的另一个实施例中,对于步骤S400,分别从各所述第一子块的第一隐空间向量分布模型上采样,基于从各所述第一子块的第一隐空间向量分布模型上的采样结果合成第一重构隐空间向量图,包括以下步骤:

[0119] 步骤S410:分别从各所述第一子块的第一隐空间向量分布模型上采样。

[0120] 假设,在步骤S100输入的目标图片为x,经过编码器提取第一隐空间向量图;经过步骤S200将第一隐空间向量图划分成多第一子块;经过步骤S300,对于其中一个第一子块而言,计算得到该第一子块的平均值为z_mean,方差为z_sigma,并构建了第一隐空间向量分布模型z。那么在本步骤中则可根据:z=z_mean+z0*z_sigma,进行重参数化采样,得到采样结果z0。对于其他第一子块均可按照上述方法进行采样。采用重参数化算法,方便在代码层面上实现采样。

[0121] 步骤S420:基于从各所述第一子块的第一隐空间向量分布模型上的采样结果,合成与所述目标图片具有相同位置分布的所述第一重构隐空间向量图。

[0122] 通过步骤S410,针对每一个第一子块的第一隐空间向量分布模型,均能够得到一个采样z0,而各个第一子块是按照长、宽两个维度进行划分的,那么各个第一子块能够保留目标图片长、宽两个维度的空间结构信息,对于在本步骤中基于从各所述第一子块的第一隐空间向量分布模型上的采样结果,合成与所述目标图片具有相同位置分布的第一重构隐空间向量图,则同样可以保留目标图片在长、宽维度上的空间结构信息。

[0123] 步骤S500:将所述第一重构隐空间向量图输入解码器,生成第一重构图片。

[0124] 其中,该第一重构图片是基于第一重构隐空间向量图进行解码生成的重构图片。

[0125] 一些实施方式中,该第一重构隐空间向量图是通过在各个第一子块的第一隐空间向量分布模型进行采样,而后重新按照各个第一子块的位置关系进行组合的,对于那么第

一重构隐空间向量和第一隐空间向量的各个位置来说，具有相同的位置关系，因此，能够准确的反映目标图片的各个位置的空间结构信息，从而基于第一重构隐空间向量图转化而成的第一重构图片也保持了目标图片原有的空间结构信息。

[0126] 如图4a所示，解码器可以由一系列的卷积神经网络组成，通过卷积神经网络能够将隐空间向量图转化成图片。解码器也可以部署在计算机、平板、手机、扫描仪等设备上，解码器将隐空间向量图转化成图片后可以借助显示器等设备进行显示。

[0127] 步骤S600：根据所述第一重构图片与所述目标图片的相似度，判断所述目标图片的真伪。

[0128] 即基于解码器重构的第一重构图片和目标图片的相似度判断目标图片是否为真实图片。

[0129] 在本申请的另一个实施例中，由于目标图片可为待检测图片、第二重构图片，因此，可分别以待检测图片、第二重构图片为例说明如何根据所述第一重构图片与所述目标图片的相似度，判断所述目标图片的真伪：

[0130] 方式一、目标图片为待检测图片

[0131] 可通过比较所述第一重构图片与所述目标图片的像素值差异，判断所述目标图片的真伪。

[0132] 假设输入的目标图片为x_unkn，最终通过解码器输出的第一重构图片为x_unkn'，x_unkn和x_unkn'分别代表了目标图片和第一重构图片的像素，那么则可以通过score1＝||x_unkn-x_unkn'||^{2}计算得到一个score1，通过score1与预设值大小进行比较就可以判断目标图片的真伪。

[0133] 方式二、目标图片为第二重构图片

[0134] 结合图4b具体来说，将所述第一重构图片输入至所述编码器，得到所述第一重构图片的第二隐空间向量图；

[0135] 基于所述第一隐空间向量图和所述第二隐空间向量图的相似度，判断所述目标图片的真伪。

[0136] 对于方式二来说，假设目标图片为x_unkn，第一隐空间空间向量图为x_f_unkn，各个第一子块分别为x_f1_unkn、x_f 2_unkn……x_fn_unkn，各个第一子块的第一隐空间向量的平均值和方差分别(z1_mean，z1_sigma)、(z2_mean，z2_sigma)……(zn_mean，zn_sigma)，各个第一子块的第一隐空间向量分布模型分比为z1、z2……zn，生成的第一重构图片为x_unkn'，那么此时则可以将第一重构图片输入步骤S100中的编码器进行步骤S100-步骤S400的操作，各个步骤具体过程不再一一重复，可以得到关于第一重构图片的第二隐空间空间向量图为x_f_unkn'，基于第二隐空间空间向量图划分的各个第二子块分别为x_f1_unkn'、x_f 2_unkn'……x_fn_unkn'，各个第二子块的第二隐空间向量的平均值和方差分别(z1_mean'，z1_sigma')、(z2_mean'，z2_sigma')……(zn_mean'，zn_sigma')，各个第二子块的第二隐空间向量分布模型分比为z1'、z2'……zn'，并分别从各个第二子块对应的第二隐空间向量分布模型上采样得到各个采样结果z0'。

[0137] 那么在本申请的一个实施例中，则可以基于所述第一隐空间向量图x_f_unkn和所述第二隐空间向量图x_f_unkn'的相似度，判断所述目标图片的真伪来进行判断目标图片的真伪。

[0138] 对于方式二来说，在本申请的另一实施例中，还可以通过比较各所述第一子块和各所述第二子块的差异，判断所述目标图片的真伪。即x_f1_unkn、x_f 2_unkn……x_fn_unkn，分别与x_f1_unkn'、x_f 2_unkn'……x_fn_unkn'之间的差异，来判断目标图片的真伪。在对比差异时，x_f1_unkn与x_f1_unkn'、x_f 2_unkn与x_f 2_unkn'、x_fn_unkn与x_fn_unkn'分别对比，各个第一子块与各个第二子块分别对比后，进行综合评估。

[0139] 对于方式二来说，在本申请的另一实施例中，可以利用所述各个第二子块和所述各个第一子块隐空间向量的方差和平均值的相似度，来表征所述各个第二子块与所述各个第一子块的相似度。即(z1_mean，z1_sigma)、(z2_mean，z2_sigma)……(zn_mean，zn_sigma)与(z1_mean'，z1_sigma')、(z2_mean'，z2_sigma')……(zn_mean'，zn_sigma')的相似度。在对比相似度时，(z1_mean，z1_sigma)与(z1_mean'，z1_sigma')、(z2_mean，z2_sigma)与(z2_mean'，z2_sigma')、(zn_mean，zn_sigma)与(z2_mean'，z2_sigma')分别对比，各个第一子块与各个第二子块的隐空间向量的平均值和方差分别对比后，进行综合评估。

[0140] 对于方式二来说，在本申请的另一个实施例中，可以基于在各所述第二子块的第二隐空间向量分布模型上的采样结果，与在各所述第一子块的第一隐空间向量分布模型上的采样结果的相似度，判断所述目标图片的真伪。即根据从各所述第一子块的第一隐空间向量分布模型的采样结果z0，与对应各所述第二子块的第二隐空间向量分布模型的采样结果z0'进行对比相似度，来判断目标图片的真伪。需要说明的是，对采样结果的对比时，基于相同位置的第一子块和第二子块分别进行对比，再结合全部对比结果进行综合评估。

[0141] 在本申请的另一实施例中，所述编码器和所述解码器经过如下训练得到：

[0142] 获取真实的训练图片；

[0143] 将所述训练图片输入所述编码器，得到所述训练图片的第三隐空间向量图；

[0144] 将所述第三隐空间向量图划分成多个第三子块；

[0145] 分别构建各所述第三子块的第三隐空间向量分布模型；

[0146] 从各所述第三子块的第三隐空间向量分布模型上采样，合成第三重构隐空间向量图；

[0147] 将所述第三重构隐空间向量图输所述编码器，生成与所述训练图片对应的第三重构图片；

[0148] 计算所述第三重构图片与对应的训练图片之间的像素重构损失和隐空间分布损失，基于所述像素重构损失和所述隐空间分布损失，调整所述解码器的模型参数和所述编码器的模型参数，直至满足预设结束条件时结束训练。

[0149] 首先获取真实的训练图片，比如对于人脸图片的检测来说，那么此时训练图片可以为人脸的真实图片，对于真实的人脸图片来说比较容易获取，可以基于一些开源的人脸数据集，或者可以从互联网获取，再或者可以对真实人脸进行拍摄获取，本申请对于真实的训练图片的来源、获取途径不做限制。

[0150] 在本申请的另一实施例中，获取的训练图片还可以有多张，以保证训练的准确性均匀性，另外，获取的多张训练图片还可以分批次进行训练，每个批次可以包括多张训练图片，以缩短训练时间，提高训练效率。

[0151] 然后将获取的训练图片输入至步骤S100中使用的编码器中，进行步骤S100-步骤

S600，各个步骤的操作方法不再一一赘述。在步骤S100中能得到关于训练图片的第三隐空间向量图，在步骤S200中，将第三隐空间向量图划分成多个第三子块，在步骤S300中，构建各个第三子块的第三隐空间向量分布模型，在步骤S400中从各个第三子块的第三隐空间向量分布模型上采样，合成第三重构隐空间向量图，在步骤S500中，基于解码器将第三重构隐空间向量图转化为第三重构图片。

[0152] 接下来计算第三重构图片与对应的训练图片之间的像素重构损失和隐空间分布损失。

[0153] 假设真实的训练图片为x，得到的第三重构图片为x'，那么第三重构图片与对应的训练图片之间的像素重构损失则可表示为$Loss_recon=||x-x'||^2$

[0154] 可以理解的是，在又一个实施例中，还可以通过余弦相似度、曼哈顿距离、明氏距离和切比雪夫距离等向量相似度的计算方式计算像素重构损失，本申请对此不做限制。

[0155] 对于第三重构图片与对应的训练图片之间的隐空间分布损失，可以基于各个第三子块的第三隐空间向量分布模型与标准正态分布之间的损失进行计算，计算过程如下：

[0156] Loss_KL＝KL(N(z_mean,z_sigma),N(0,I))，

[0157] 其中(z_mean,z_sigma)为各个第三子块的隐空间向量的平均值和方差，N(0,I)为标准正态分布，KL为各个第三隐空间向量分布模型与标准正态分布之间的散度，Loss_KL即为第三重构图片与对应的训练图片之间的隐空间分布损失。

[0158] 可以理解的是，在又一个实施例中，还可以通过JS散度、Wasserstein距离、海林格距离等概率分布的相似度计算方式计算各个第三子块的第三隐空间向量分布模型与标准正态分布之间的损失，本申请对此不做限制。

[0159] 然后基于Loss_KL和Loss_recon，利用随机梯度下降算法对所述解码器的模型参数和所述编码器的模型参数进行更新，直至所述编码器的模型参数和所述解码器的模型参数收敛。

[0160] 在本申请的一实施例中，可以利用公式1中的loss表征第三重构图片与对应的训练图片之间的隐空间分布损失和重构损失的整体损失。

[0161] Loss＝Loss_KL+λ*Loss_recon公式1

[0162] 其中，λ可取经验值1，然后利用随机梯度下降算法对所述解码器的模型参数和所述编码器的模型参数进行更新，直至整体损失不再下降，此时就得到了训练后的解码器和编码器。

[0163] 为便于理解，下面以人脸识别应用场景为例，对本申请实施例中的伪造图片检测方法进行说明。如图12所示，为利用本申请实施例中解码器和编码器的训练方法所得到的人脸识别模型，图12左侧为输入真实人脸图片得到的重构人脸图片，图12右侧为输入伪造人脸图片得到重构人脸图片，可以清楚的看出基于真实人脸图片的重构图片改变较小，而基于伪造人脸图片得到的重构图片改变较大，因此利用经过上述训练后的编码器和解码器所构成的人脸图片识别模型能够清楚的分辨输入的人脸图片的真假。

[0164] 接下来，通过如下表1中的实验，对本申请实施例中的深度伪造图片检测方法的可靠性进行验证。

[0165] 1)准备数据集，包括两种不同源的数据集，记为数据集1和数据集2；

[0166] 2)数据集划分，划分数据集1为训练集和测试集，记为train1和test1，划分数据集

2为训练集和测试集，记为train2和test2。

[0167] 3)验证过程，对照当前有监督的训练模型(例如xception)和利用本申请实施例中所阐述的对编码器和解码器进行训练的无监督的训练模型(根据打分方式不同，分为无监督模型1和无监督模型2，无监督模型1是计算像素级别的差异，本实验选取的阈值为0.5，无监督模型2计算的隐空间向量的差异，本实验选取的阈值是0.5)，分别设计两组实验，两个对照模型在train1上训练，在test1和test2上测试，两个对照模型在train2上训练，在test1和test2上测试。实验结果如下表1所示。训练集train1和测试集test1同源，训练集train2和测试集test2也同源，不难看出，对于同源数据集，有监督的方式略优于本申请的无监督的方法。

[0168] 表1

[0169]

	训练集	test1	test2
有监督二分类模型	train1	93%	53%
无监督模型 1		90%	83%
无监督模型 2		91.2%	85%
有监督二分类模型	train2	49%	96%
无监督模型 1		88%	91%
无监督模型 2		86%	92%

[0170] 在实验结果对比中，训练集train1和测试集test2不同源，训练集train2和测试集test1也不同源，对应的实验结果可以看出，本申请阐述的对于解码器和编码器的无监督的训练方法有更好的泛化性能。

[0171] 相较于现有技术中的变分自动编码器对目标图片的隐空间向量图进行一维度摊平操作，然后再基于摊平后的一维度向量图进行构建隐空间向量分布模型，在摊平操作时破坏了目标图片自身的空间结构信息，故而构建的隐空间向量分布模型并不准确，那么基于此隐空间向量分布模型进行取样重构的图片也不具有目标图片的空间结构信息，基于此与目标图片进行对比判断真伪的准确率较低。而本申请实施例中，是通过将目标图片的隐空间向量图划分成多个第一子块，然后再基于划分后的多个第一子块分别构建隐空间向量分布模型，在划分多个第一子块时，各个第一子块能够保留目标图片自身的空间结构信息，故而基于各个第一子块构建的各个隐空间向量分布模型也能够保留目标图片自身的空间结构信息，基于从各个隐空间向量分布模型的采样结果，重构的第一隐空间向量分布图也能够保留目标图片的空间结构信息，那么基于第一重构隐空间向量图重构的第一重构图片也与目标图片具有相同的空间结构信息，基于此进行对比相似度判断目标图片的真伪的准确率较高。

[0172] 示例性装置

[0173] 图2至图12中任一项所对应的实施例中所提及的任一技术特征也同样适用于本申请实施例中的图13-图15所对应的实施例，后续类似之处不再赘述。在介绍了本申请实施例示例性实施方式的伪造图片检测方法之后，接下来对本申请实施例中实施上述伪造图片检测方法的伪造图片检测装置进行详细阐述。

[0174] 如图13所示的一种伪造图片检测装置500，其可应用于图片检测领域，具体用于检测伪造图片等操作。本申请实施例中的伪造图片检测装置能够实现对应于上述图2-图12中

任一项所对应的实施例中所执行的伪造图片检测方法的步骤。该伪造图片检测装置500实现的功能可以通过硬件实现，也可以通过硬件执行相应的软件实现。硬件或软件包括一个或多个与上述功能相对应的模块，所述模块可以是软件和/或硬件。所述伪造图片装置可包括输入输出模块100、编码器200、处理模块300和解码器400，所述输入输出模块100、编码器200、处理模块300和解码器400的功能实现可参考图2-图12中任一项所对应的实施例中所执行的操作，此处不作赘述。例如，所述处理模块300可用于对第一隐空间向量分布模型采样，合成第一重构隐空间向量图、构建各所述第一子块的第一隐空间向量分布模型等操作。

[0175] 一些实施方式中，所述输入输出模块100可用于向编码器200输入目标图片；

[0176] 所述编码器200可用于对自所述输入输出模块100输入的所述目标图片进行编码处理，得到所述目标图片的第一隐空间向量图，将所述第一隐空间向量图划分成多个第一子块，构建各所述第一子块的第一隐空间向量分布模型；

[0177] 所述处理模块300可用于从编码器200构建的各所述第一子块的第一隐空间向量分布模型上采样，合成第一重构隐空间向量图；

[0178] 所述解码器400可用于将所述处理模块300合成的第一重构隐空间向量图解码，生成与所述目标图片对应的第一重构图片；

[0179] 所述处理模块300可根据所述解码器400解码生成的所述第一重构图片与所述目标图片的相似度，判断所述目标图片的真伪。

[0180] 在本申请的另一个实施例中，所述编码器200还用于：基于长、宽、通道中的至少一个维度将所述第一隐空间向量图划分成所述多个第一子块。

[0181] 在本申请的另一个实施例中，所述编码器200用于：基于从所述目标图片提取的所述第一隐空间向量图，分别计算各所述多个第一子块的隐空间向量的平均值和方差；

[0182] 基于各所述多个第一子块的隐空间向量的平均值和方差，分别构建各所述第一子块的第一隐空间向量分布模型。

[0183] 在本申请的另一个实施例中，所述处理模块300用于从所述编码器200构建的各所述第一子块的第一隐空间向量分布模型上采样，合成与所述目标图片具有相同位置分布的所述第一重构隐空间向量图。

[0184] 在本申请的另一个实施例中，所述处理模块300用于：

[0185] 将所述解码器400生成的第一重构图片输入至所述编码器200，所述编码器200对所述第一重构图片进行提取第二隐空间向量图；

[0186] 所述处理模块300通过比较所述第一隐空间向量图和所述第二隐空间向量图的相似度，判断所述目标图片的真伪。

[0187] 在本申请的另一个实施例中，所述编码器200将所述第二隐空间向量图划分成与所述第一隐空间向量图相同的多个第二子块，所述处理模块300通过比较各个所述第二子块与各个所述第一子块的相似度，判断所述目标图片的真伪。

[0188] 在本申请的另一个实施例中，所述编码器200分别计算各个所述第二子块以及各个所述第一子块隐空间向量的方差和平均值，所述处理模块300基于各个所述第二子块以及各个所述第一子块隐空间向量的方差和平均值的相似度，判断所述目标图片的真伪。

[0189] 在本申请的另一个实施例中，所述编码器200基于所述第二隐空间向量图，构建各个所述第二子块的第二隐空间向量分布模型；

[0190] 所述处理模块300从各所述第二子块的第二隐空间向量分布模型上采样；

[0191] 所述处理模块300基于从各所述第二子块的第二隐空间向量分布模型上的采样结果，以及所述处理模块300在各所述第一子块的第一隐空间向量分布模型上的采样结果的相似度，判断所述目标图片的真伪。

[0192] 在本申请的另一个实施例中，所述处理模块300通过比较所述解码器400生成的所述第一重构图片与所述目标图片的像素值相似度，判断所述目标图片的真伪。

[0193] 在本申请的另一个实施例中，所述解码器400和所述编码器200经过如下训练得到：

[0194] 获取真实的训练图片；

[0195] 通过所述输入输出模块100将所述训练图片输入所述编码器200，得到所述训练图片的第三隐空间向量图；

[0196] 通过编码器200将所述第三隐空间向量图划分成多个第三子块；

[0197] 通过编码器200分别构建各所述第三子块的第三隐空间向量分布模型；

[0198] 处理模块300从各所述第三子块的第三隐空间向量分布模型上采样，合成第三重构隐空间向量图；

[0199] 处理模块300将所述第三重构隐空间向量图输入所述解码器400，所述解码器400对所述第三重构进空间向量图进行解码，生成与所述训练图片对应的第三重构图片；

[0200] 处理模块300计算所述第三重构图片与对应的训练图片之间的像素重构损失和隐空间分布损失，基于所述像素重构损失和所述隐空间分布损失，调整所述编码器200的模型参数和所述解码器400的模型参数，直至满足预设结束条件时结束训练。

[0201] 在本申请的另一个实施例中，所述处理模块300还用于：

[0202] 基于所述编码器200构建的各所述第三子块的第三隐空间向量分布模型和标准高斯分布计算所述隐空间分布损失；

[0203] 基于所述解码器400解码生成的第三重构图片与对应的训练图片计算所述像素损失；

[0204] 基于所述隐空间分布损失和所述像素损失，利用随机梯度下降算法对所述编码器200的模型参数和所述解码器400的模型参数进行更新，直至所述解码器400的模型参数和所述编码器200的模型参数收敛。

[0205] 相较于现有技术中的变分自动编码器对目标图片的隐空间向量图进行一维度摊平操作，然后再基于摊平后的一维度向量图进行构建隐空间向量分布模型，在摊平操作时破坏了目标图片自身的空间结构信息，故而构建的隐空间向量分布模型并不准确，那么基于此隐空间向量分布模型进行取样重构的图片也不具有目标图片的空间结构信息，基于此与目标图片进行对比判断真伪的准确率较低。而本申请实施例中，则是处理模块300通过将目标图片的隐空间向量图划分成多个第一子块，然后再基于划分后的多个第一子块分别构建隐空间向量分布模型，在划分多个第一子块时，各个第一子块能够保留目标图片自身的空间结构信息，故而基于各个第一子块构建的各个隐空间向量分布模型也能够保留目标图片自身的空间结构信息，基于从各个隐空间向量分布模型的采样结果，重构的第一隐空间向量分布图也能够保留目标图片的空间结构信息，那么基于第一重构隐空间向量图重构的第一重构图片也与目标图片具有相同的空间结构信息，基于此进行对比相似度判断目标图

片的真伪的准确率较高。

[0206] 示例性计算机可读存储介质

[0207] 在介绍了本申请实施例示例性实施方式的方法、装置之后，接下来，本申请实施例示例性实施方式的计算机可读存储介质进行说明。

[0208] 在本申请实施例中，计算机可读存储介质为光盘，其上存储有计算机程序(即程序产品)，所述计算机程序在被处理器运行时，会实现上述方法实施方式中所记载的各步骤，例如:将目标图片输入编码器，得到所述目标图片的第一隐空间向量图;将所述第一隐空间向量图划分成多个第一子块;分别构建所述多个第一子块中各所述第一子块的第一隐空间向量分布模型;分别从各所述第一子块的第一隐空间向量分布模型上采样，基于从各所述第一子块的第一隐空间向量分布模型上的采样结果合成第一重构隐空间向量图;将所述第一重构隐空间向量图输入解码器，生成第一重构图片;根据所述第一重构图片与所述目标图片的相似度，判断所述目标图片的真伪。各步骤的具体实现方式在此不再重复说明。

[0209] 需要说明的是，所述计算机可读存储介质的例子还可以包括，但不限于相变内存(PRAM)、静态随机存取存储器(SRAM)、动态随机存取存储器(DRAM)、其他类型的随机存取存储器(RAM)、只读存储器(ROM)、电可擦除可编程只读存储器(EEPROM)、快闪记忆体或其他光学、磁性存储介质，在此不再一一赘述。

[0210] 示例性处理设备

[0211] 上面从模块化功能实体的角度对本申请实施例中的伪造图片检测装置500进行了描述，下面从硬件处理的角度分别对本申请实施例中的执行伪造图片检测方法的服务器、终端进行描述。需要说明的是，在本申请实施例图14所示的实施例中的输入输出模块100对应的实体设备可以为输入/输出单元、收发器、射频电路、通信模块和输入/输出(I/O)接口等，编码器200、解码器400和处理模块300对应的实体设备可以为处理器。图13所示的伪造图片检测装置500可以具有如图14所示的结构，当图13所示的伪造图片检测装置500具有如图14所示的结构时，图14中的处理器和I/O接口能够实现前述对应该装置的装置实施例提供的处理模块300、编码器200、解码器400和输入输出模块100相同或相似的功能，图14中的存储器存储处理器执行上述伪造图片检测方法时需要调用的计算机程序。

[0212] 具体来说，图14示出了适于用来实现本申请实施例实施方式的示例性计算设备80的框图，该计算设备80可以是计算机系统或服务器。图14显示的计算设备80仅仅是一个示例，不应对本申请实施例的功能和使用范围带来任何限制。

[0213] 如图14所示，计算设备80的组件可以包括但不限于:一个或者多个处理器或者处理单元801，系统存储器802，连接不同系统组件(包括系统存储器802和处理单元801)的总线803。

[0214] 计算设备80典型地包括多种计算机系统可读介质。这些介质可以是任何能够被计算设备80访问的可用介质，包括易失性和非易失性介质，可移动的和不可移动的介质。

[0215] 系统存储器802可以包括易失性存储器形式的计算机系统可读介质，例如随机存取存储器(RAM)8021和/或高速缓存存储器8022。计算设备70可以进一步包括其它可移动/不可移动的、易失性/非易失性计算机系统存储介质。仅作为举例，ROM8023可以用于读写不可移动的、非易失性磁介质(图14中未显示，通常称为“硬盘驱动器”)。尽管未在图14中示出，可以提供用于对可移动非易失性磁盘(例如“软盘”)读写的磁盘驱动器，以及对可移动

非易失性光盘(例如CD-ROM,DVD-ROM或者其它光介质)读写的光盘驱动器。在这些情况下,每个驱动器可以通过一个或者多个数据介质接口与总线803相连。系统存储器802中可以包括至少一个程序产品,该程序产品具有一组(例如至少一个)程序模块,这些程序模块被配置以执行本申请各实施例的功能。

[0216] 具有一组(至少一个)程序模块8024的程序/实用工具8025,可以存储在例如系统存储器802中,且这样的程序模块8024包括但不限于:操作系统、一个或者多个应用程序、其它程序模块以及程序数据,这些示例中的每一个或某种组合中可能包括网络环境的实现。程序模块8024通常执行本申请所描述的实施例中的功能和/或方法。

[0217] 计算设备80也可以与一个或多个外部设备804(如键盘、指向设备、显示器等)通信。这种通信可以通过输入/输出(I/O)接口进行。并且,计算设备80还可以通过网络适配器806与一个或者多个网络(例如局域网(LAN),广域网(WAN)和/或公共网络,例如因特网)通信。如图14所示,网络适配器806通过总线803与计算设备80的其它模块(如处理单元801等)通信。应当明白,尽管图14中未示出,可以结合计算设备80使用其它硬件和/或软件模块。

[0218] 处理单元801通过运行存储在系统存储器802中的程序,从而执行各种功能应用以及数据处理,例如,将目标图片输入编码器,得到所述目标图片的第一隐空间向量图;将所述第一隐空间向量图划分成多个第一子块;分别构建所述多个第一子块中各所述第一子块的第一隐空间向量分布模型;分别从各所述第一子块的第一隐空间向量分布模型上采样,基于从各所述第一子块的第一隐空间向量分布模型上的采样结果合成第一重构隐空间向量图;将所述第一重构隐空间向量图输入解码器,生成第一重构图片;根据所述第一重构图片与所述目标图片的相似度,判断所述目标图片的真伪。各步骤的具体实现方式在此不再重复说明。应当注意,尽管在上文详细描述中提及了伪造图片检测装置的若干单元/模块或子单元/子模块,但是这种划分仅仅是示例性的并非强制性的。实际上,根据本申请实施例的实施方式,上文描述的两个或更多单元/模块的特征和功能可以在一个单元/模块中具体化。反之,上文描述的一个单元/模块的特征和功能可以进一步划分为由多个单元/模块来具体化。

[0219] 图15是本申请实施例提供的一种服务器结构示意图,该服务器1100可因配置或性能不同而产生比较大的差异,可以包括一个或一个以上中央处理器(英文全称:central processing units,英文简称:CPU)1122(例如,一个或一个以上处理器)和存储器1132,一个或一个以上存储应用程序1142或数据1144的存储介质1130(例如一个或一个以上海量存储设备)。其中,存储器1132和存储介质1130可以是短暂存储或持久存储。存储在存储介质1130的程序可以包括一个或一个以上模块(图示没标出),每个模块可以包括对服务器中的一系列指令操作。更进一步地,中央处理器1122可以设置为与存储介质1130通信,在服务器1100上执行存储介质1130中的一系列指令操作。

[0220] 服务器1110还可以包括一个或一个以上电源1120,一个或一个以上有线或无线网络接口1150,一个或一个以上输入输出接口1158,和/或,一个或一个以上操作系统1141,例如Windows Server,Mac OS X,Unix,Linux,FreeBSD等等。

[0221] 上述实施例中由服务器所执行的步骤可以基于该图15所示的服务器1100的结构。例如,例如上述实施例中由图13所示的伪造图片检测装置500所执行的步骤可以基于该图15所示的服务器结构。例如,所述中央处理器1122通过调用存储器1132中的指令,执行以下

操作：

[0222] 通过输入输出接口1158向应用程序1142中的编码器程序输入目标图片；

[0223] 编码器程序，对自输入输出接口1158输入的所述目标图片进行编码处理，得到所述目标图片的第一隐空间向量图，将所述第一隐空间向量图划分成多个第一子块，构建各所述第一子块的第一隐空间向量分布模型；

[0224] 中央处理器1122从编码器程序构建的各所述第一子块的第一隐空间向量分布模型上采样，合成第一重构隐空间向量图；

[0225] 解码器程序将所述中央处理器1122合成的第一重构隐空间向量图解码，生成与所述目标图片对应的第一重构图片；

[0226] 中央处理器1122通过比较所述解码器程序解码生成的所述第一重构图片与所述目标图片的相似度，判断所述目标图片的真伪。

[0227] 在上述实施例中，对各个实施例的描述都各有侧重，某个实施例中没有详述的部分，可以参见其他实施例的相关描述。

[0228] 所属领域的技术人员可以清楚地了解到，为描述的方便和简洁，上述描述的系统，装置和模块的具体工作过程，可以参考前述方法实施例中的对应过程，在此不再赘述。

[0229] 在本申请实施例所提供的几个实施例中，应该理解到，所揭露的系统，装置和方法，可以通过其它的方式实现。例如，以上所描述的装置实施例仅仅是示意性的，例如，所述模块的划分，仅仅为一种逻辑功能划分，实际实现时可以有另外的划分方式，例如多个模块或组件可以结合或者可以集成到另一个系统，或一些特征可以忽略，或不执行。另一点，所显示或讨论的相互之间的耦合或直接耦合或通信连接可以是通过一些接口，装置或模块的间接耦合或通信连接，可以是电性，机械或其它的形式。

[0230] 所述作为分离部件说明的模块可以是或者也可以不是物理上分开的，作为模块显示的部件可以是或者也可以不是物理模块，即可以位于一个地方，或者也可以分布到多个网络模块上。可以根据实际的需要选择其中的部分或者全部模块来实现本实施例方案的目的。

[0231] 另外，在本申请实施例各个实施例中的各功能模块可以集成在一个处理模块中，也可以是各个模块单独物理存在，也可以两个或两个以上模块集成在一个模块中。上述集成的模块既可以采用硬件的形式实现，也可以采用软件功能模块的形式实现。所述集成的模块如果以软件功能模块的形式实现并作为独立的产品销售或使用时，可以存储在一个计算机可读取存储介质中。

[0232] 在上述实施例中，可以全部或部分地通过软件、硬件、固件或者其任意组合来实现。当使用软件实现时，可以全部或部分地以计算机程序产品的形式实现。

[0233] 所述计算机程序产品包括一个或多个计算机指令。在计算机上加载和执行所述计算机计算机程序时，全部或部分地产生按照本申请实施例所述的流程或功能。所述计算机可以是通用计算机、专用计算机、计算机网络、或者其他可编程装置。所述计算机指令可以存储在计算机可读存储介质中，或者从一个计算机可读存储介质向另一计算机可读存储介质传输，例如，所述计算机指令可以从一个网站站点、计算机、服务器或数据中心通过有线(例如同轴电缆、光纤、数字用户线(DSL))或无线(例如红外、无线、微波等)方式向另一个网站站点、计算机、服务器或数据中心进行传输。所述计算机可读存储介质可以是计算机能够

存储的任何可用介质或者是包含一个或多个可用介质集成的服务器、数据中心等数据存储设备。所述可用介质可以是磁性介质,(例如,软盘、硬盘、磁带)、光介质(例如,DVD)、或者半导体介质(例如固态硬盘Solid State Disk(SSD))等。

[0234] 以上对本申请实施例所提供的技术方案进行了详细介绍,本申请实施例中应用了具体个例对本申请实施例的原理及实施方式进行了阐述,以上实施例的说明只是用于帮助理解本申请实施例的方法及其核心思想;同时,对于本领域的一般技术人员,依据本申请实施例的思想,在具体实施方式及应用范围上均会有改变之处,综上所述,本说明书内容不应理解为对本申请实施例的限制。

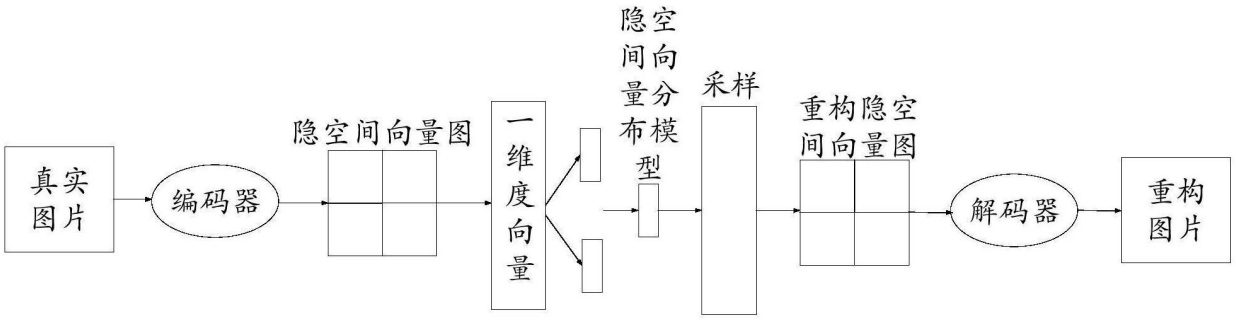

真实图片
编码器
隐空间向量图
一维度向量
隐空间向量分布模型
采样
重构隐空间向量图
解码器
重构图片

图1

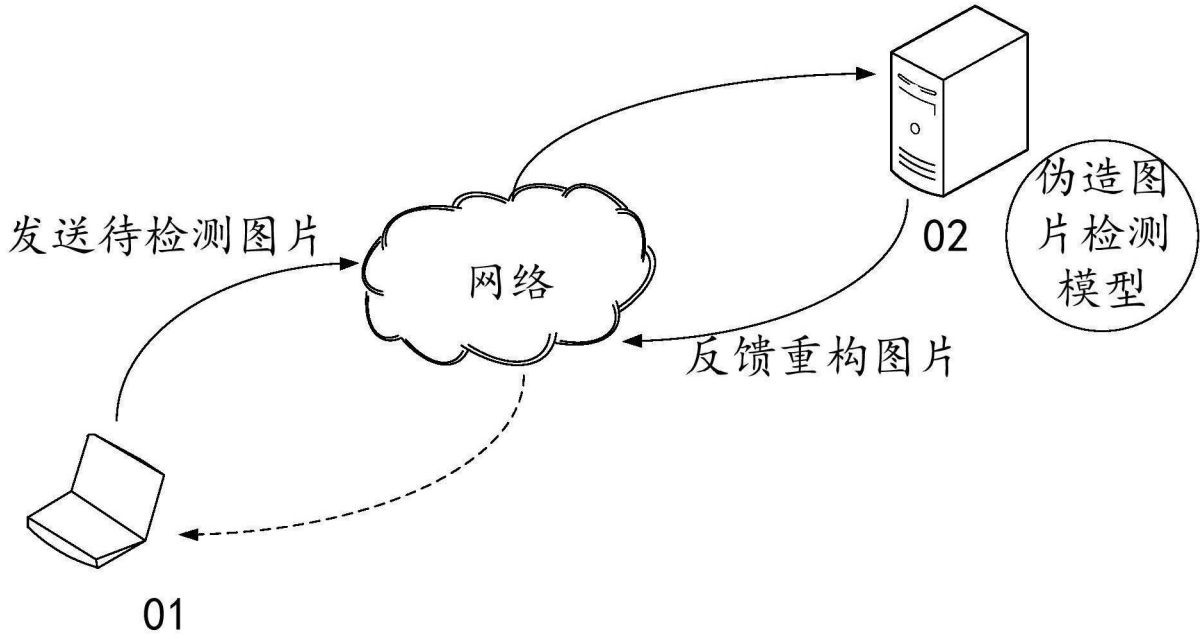

发送待检测图片
网络
02
伪造图片检测模型
反馈重构图片
01

图2a

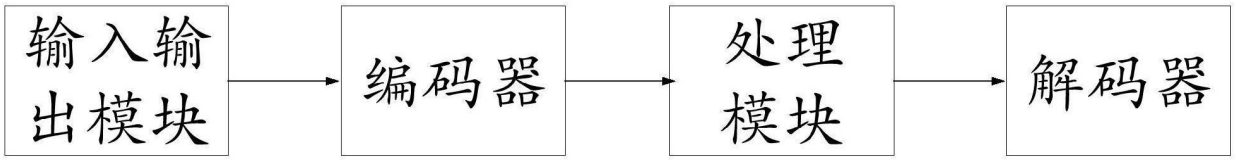

输入输出模块
编码器
处理模块
解码器

图2b

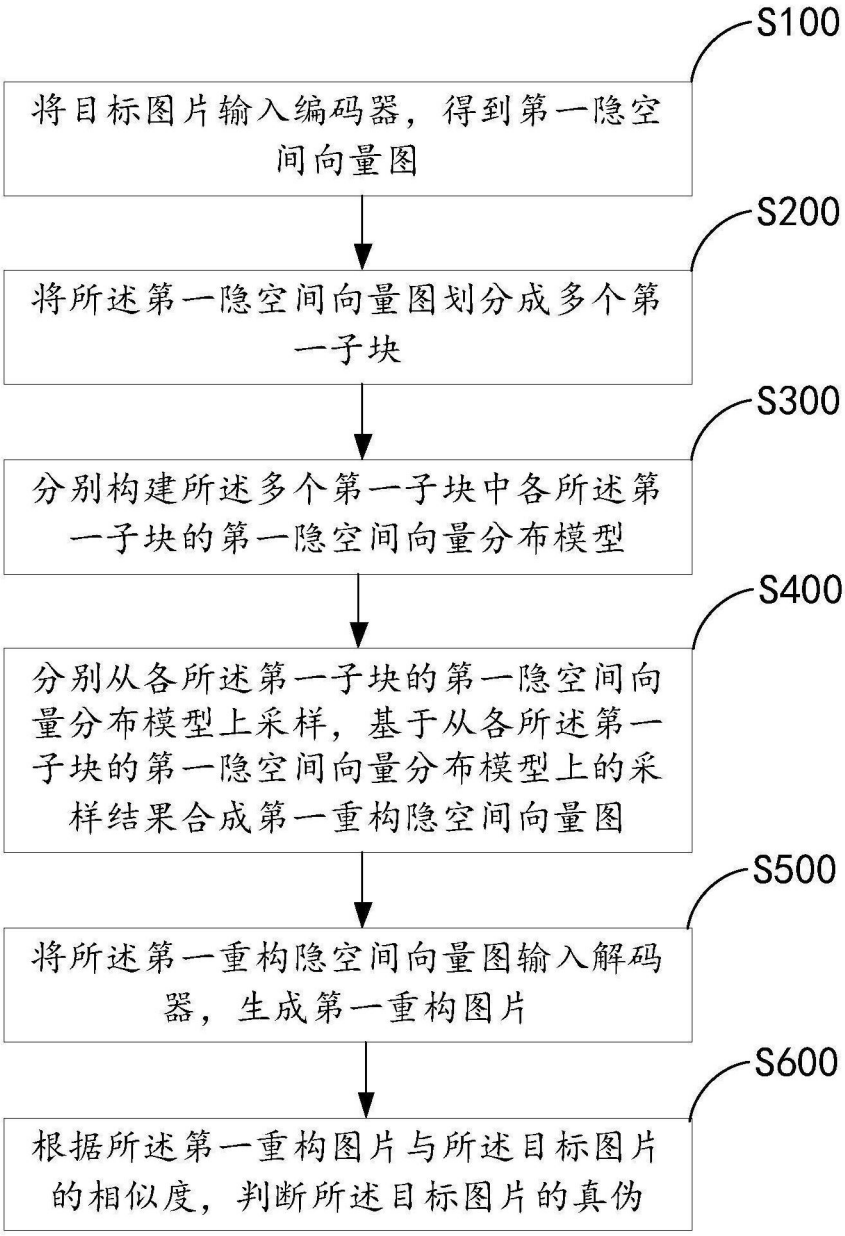

S100
将目标图片输入编码器，得到第一隐空间向量图
S200
将所述第一隐空间向量图划分成多个第一子块
S300
分别构建所述多个第一子块中各所述第一子块的第一隐空间向量分布模型
S400
分别从各所述第一子块的第一隐空间向量分布模型上采样，基于从各所述第一子块的第一隐空间向量分布模型上的采样结果合成第一重构隐空间向量图
S500
将所述第一重构隐空间向量图输入解码器，生成第一重构图片
S600
根据所述第一重构图片与所述目标图片的相似度，判断所述目标图片的真伪

图3

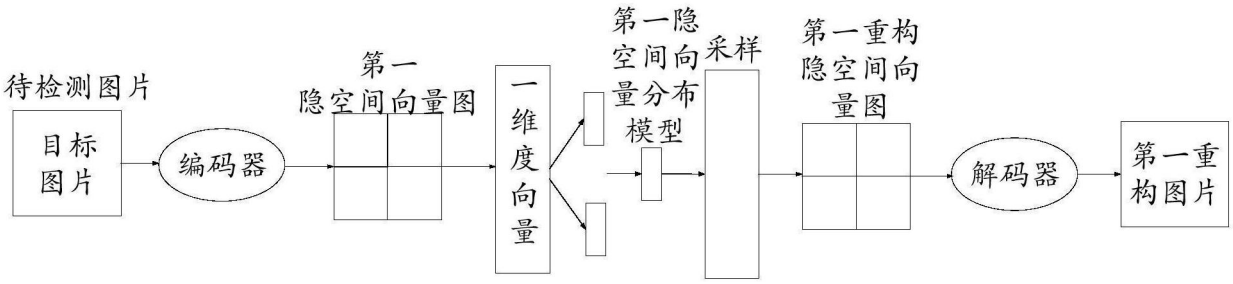

待检测图片
目标图片
编码器
第一隐空间向量图
一维度向量
第一隐空间向量分布模型
采样
第一重构隐空间向量图
解码器
第一重构图片

图4a

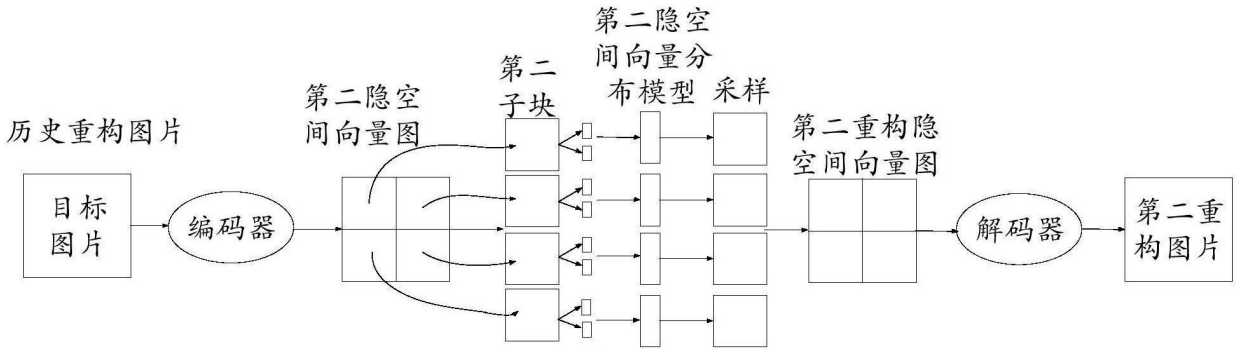

历史重构图片
目标图片
编码器
第二隐空间向量图
第二子块
第二隐空间向量分布模型
采样
第二重构隐空间向量图
解码器
第二重构图片

图4b

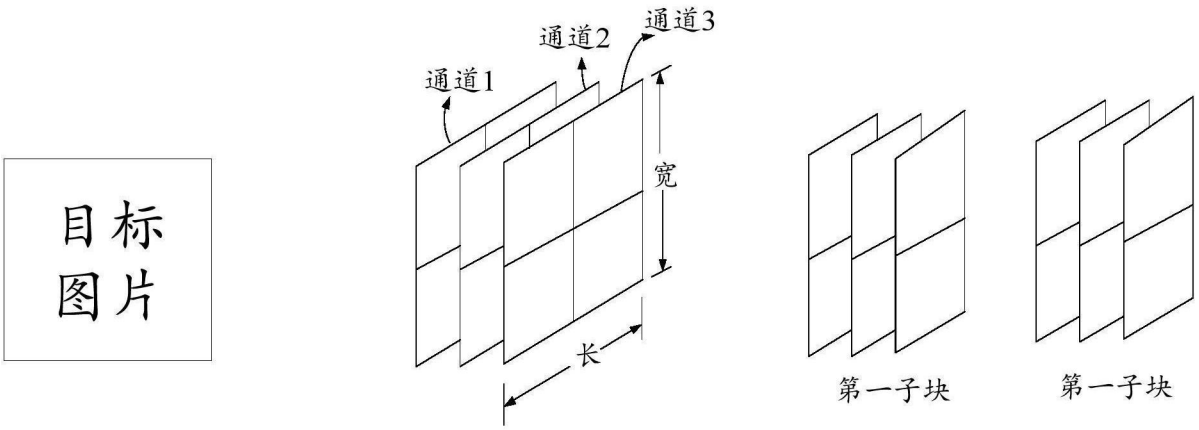

目标图片
通道1
通道2
通道3
宽
长
第一子块
第一子块

图5

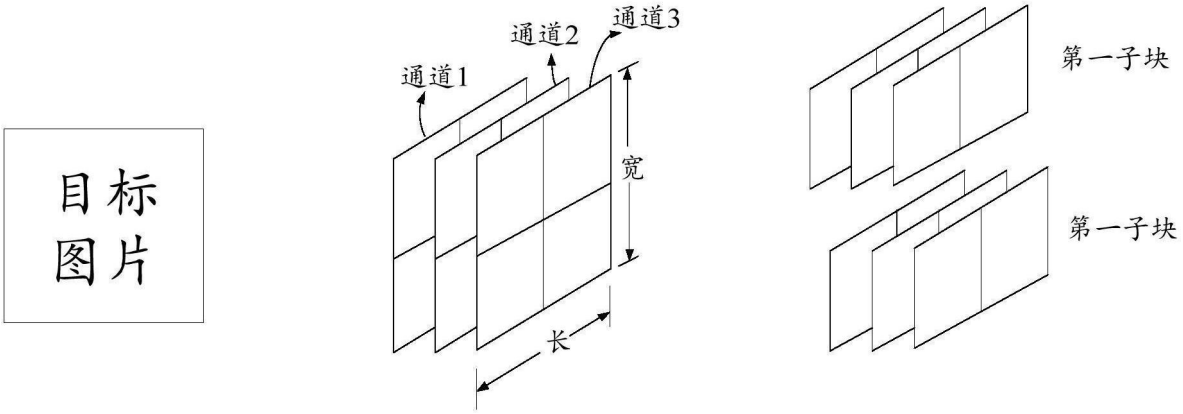

目标图片
通道1
通道2
通道3
宽
长
第一子块
第一子块

图6

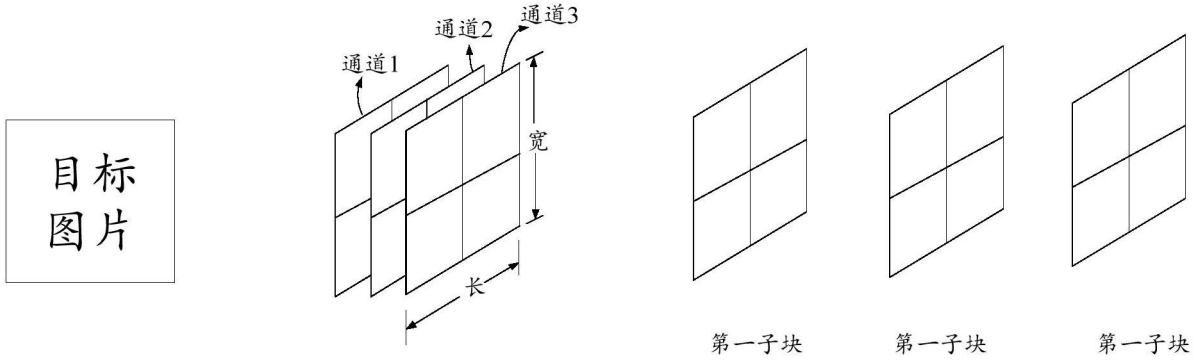

目标图片
通道1
通道2
通道3
宽
长
第一子块
第一子块
第一子块

图7

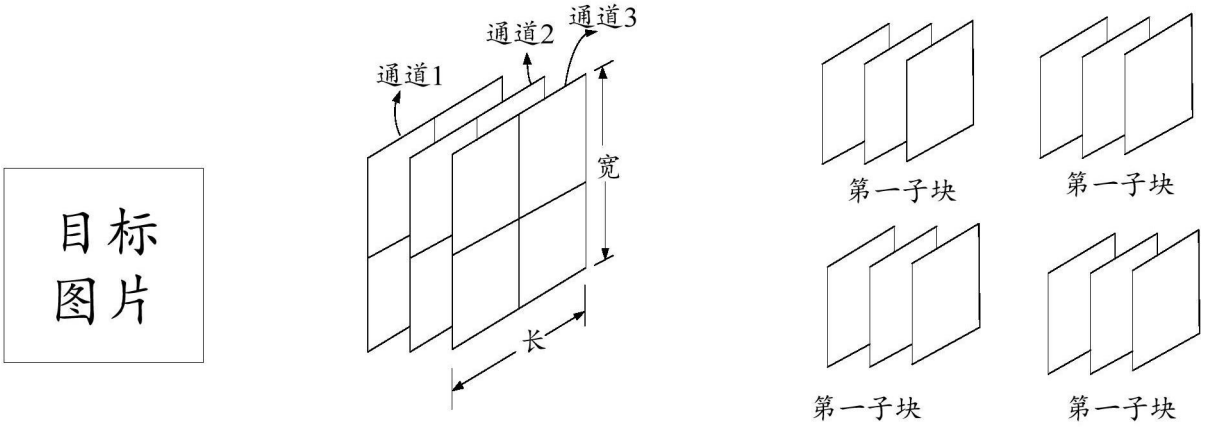

目标图片
通道1
通道2
通道3
宽
长
第一子块
第一子块
第一子块
第一子块

图8

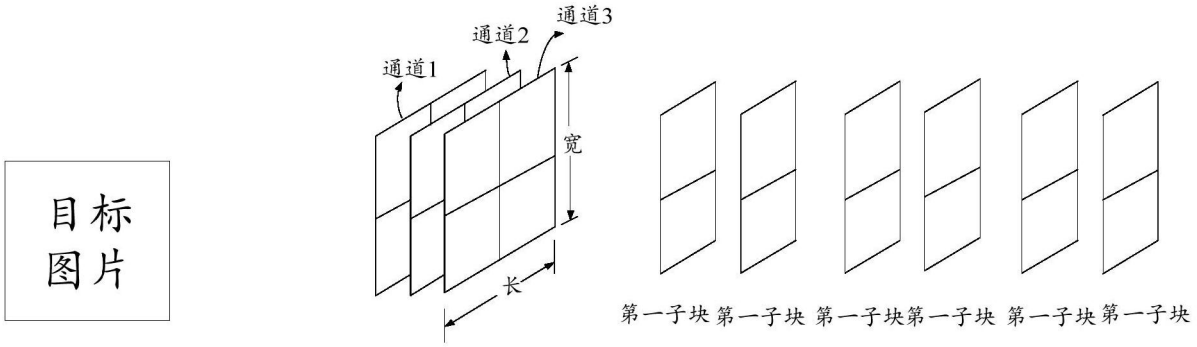

目标图片
通道1
通道2
通道3
宽
长
第一子块 第一子块 第一子块 第一子块 第一子块 第一子块

图9

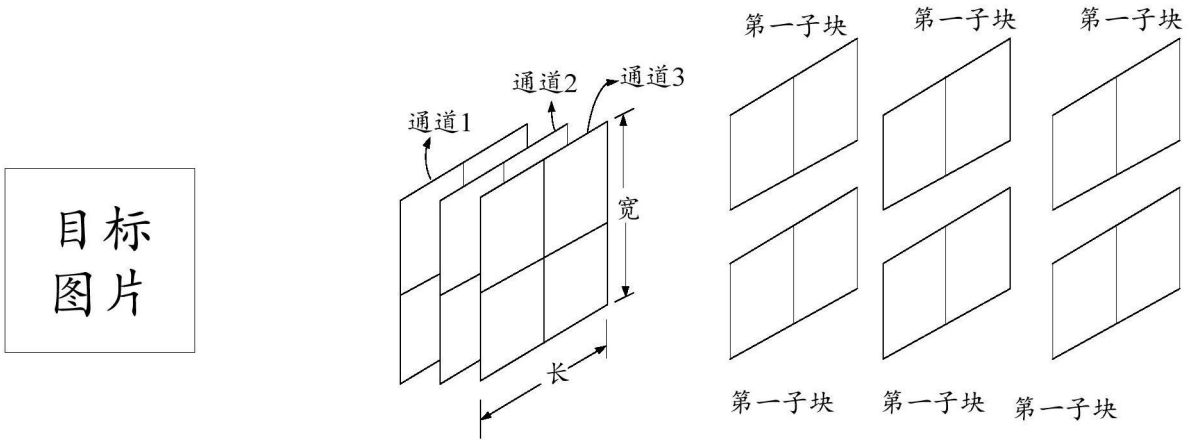

目标图片
通道1
通道2
通道3
宽
长
第一子块
第一子块
第一子块
第一子块
第一子块
第一子块

图10

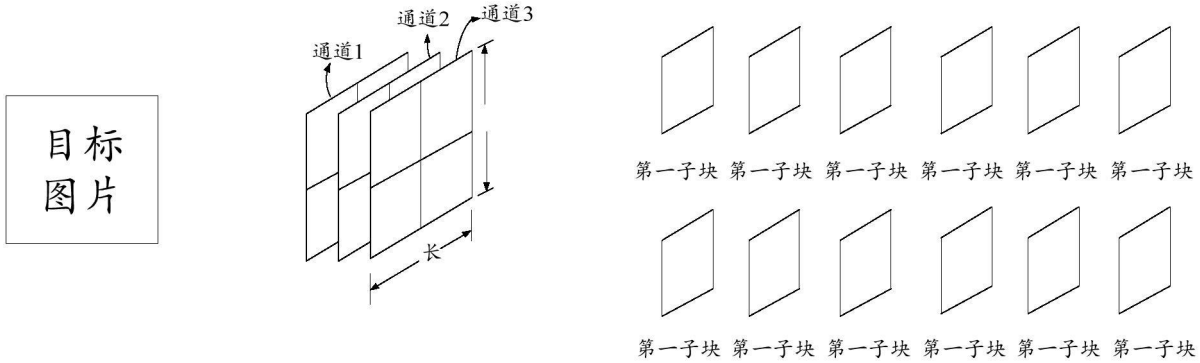

目标图片
通道1
通道2
通道3
长
第一子块 第一子块 第一子块 第一子块 第一子块 第一子块
第一子块 第一子块 第一子块 第一子块 第一子块 第一子块

图11

图12

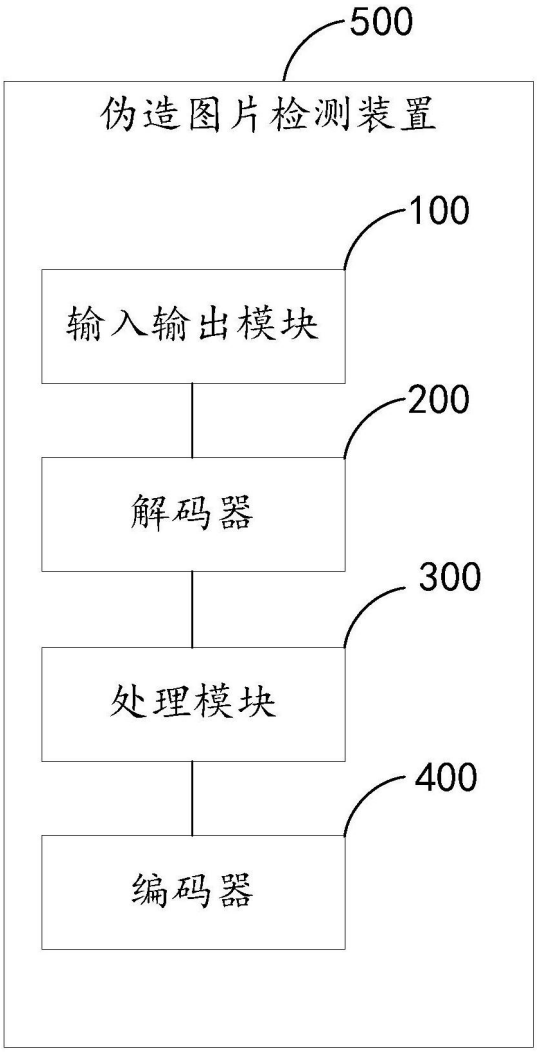


图13

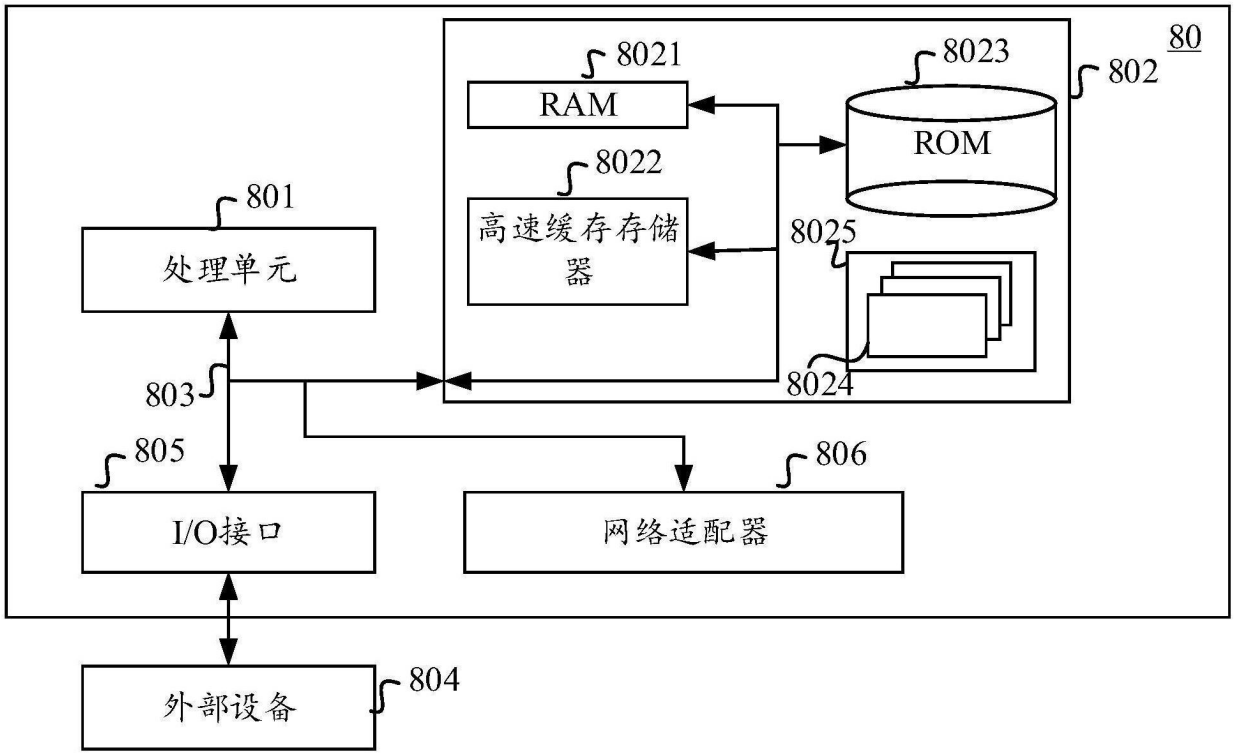

80
8021
RAM
8023
ROM
802
8022
高速缓存存储器
8025
801
处理单元
803
8024
805
806
I/O接口
网络适配器
804
外部设备

图14

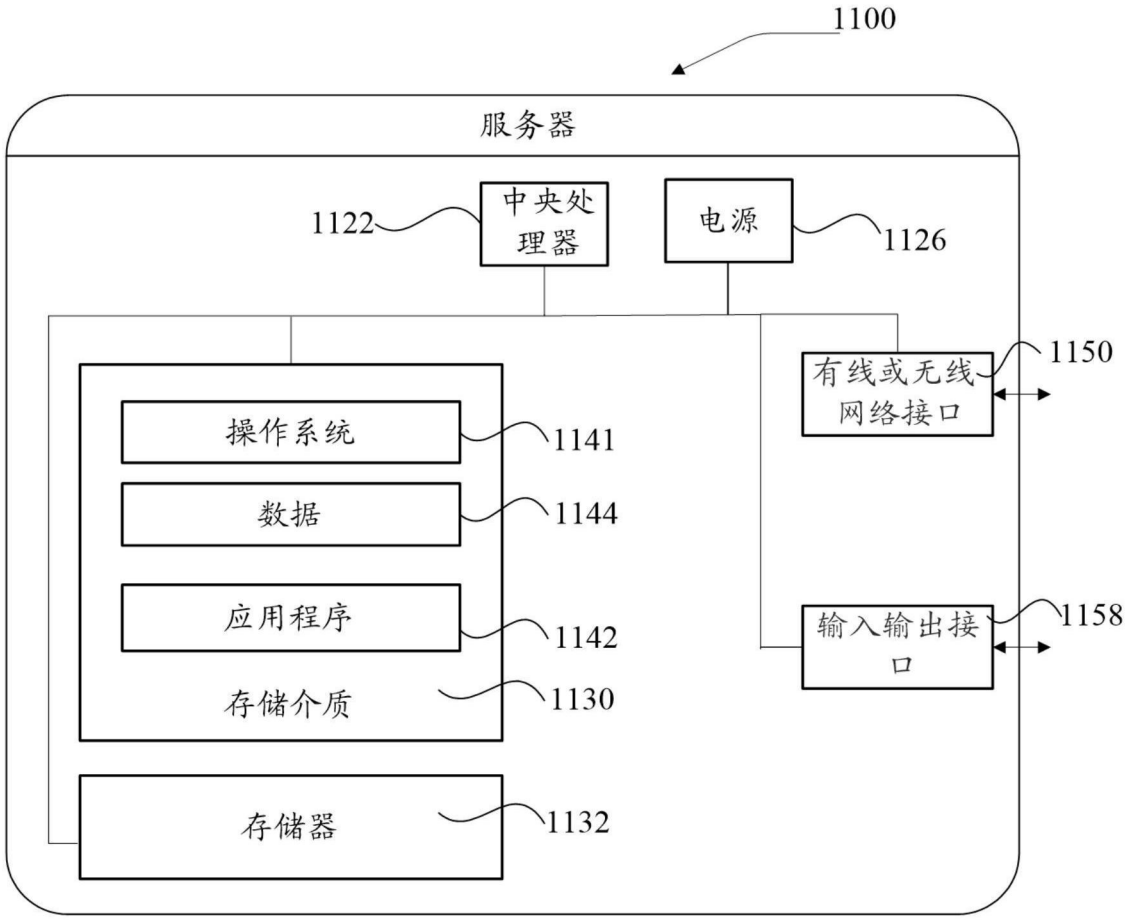

1100
服务器
1122
中央处理器
电源
1126
有线或无线网络接口
1150
操作系统
1141
数据
1144
应用程序
1142
存储介质
1130
输入输出接口
1158
存储器
1132

图15