(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 113052203 A
(43)申请公布日 2021.06.29

(21)申请号 202110181592.1

(22)申请日 2021.02.09

(71)申请人 哈尔滨工业大学(深圳)(哈尔滨工业大学深圳科技创新研究院)
地址 518055 广东省深圳市南山区桃源街道深圳大学城哈尔滨工业大学校区

(72)发明人 廖清 曾子辉 柴合言 苏伟俊 刘洋 蒋琳 王轩

(74)专利代理机构 广州三环专利商标代理有限公司 44202
代理人 郭浩辉 颜希文

(51)Int.Cl.
G06K 9/62(2006.01)
G06N 3/04(2006.01)
G06N 3/08(2006.01)

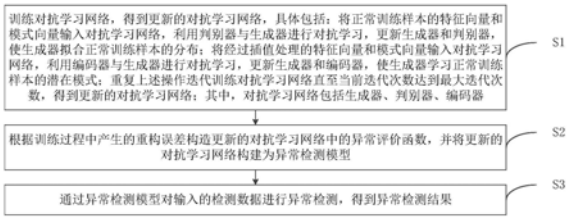
权利要求书3页 说明书11页 附图2页

(54)发明名称

一种面向多种类数据的异常检测方法及装置

(57)摘要

本发明公开了一种面向多种类数据的异常检测方法及装置。所述面向多种类数据的异常检测方法通过训练对抗学习网络,使对抗学习网络中的生成器拟合正常训练样本的分布以及学习正常训练样本的潜在模式,得到更新的对抗学习网络,根据训练过程中产生的重构误差构造更新的对抗学习网络中的异常评价函数,并将更新的对抗学习网络构建为异常检测模型,以利用异常检测模型对输入的检测数据进行异常检测,得到异常检测结果。本发明基于传统生成对抗学习模型的异常检测方法,通过引入模式分类器的思想,有效解决了检测数据与正常数据分布相近时异常检测难的问题,进一步提高了异常检测的准确性。



CN 113052203 A

1.一种面向多种类数据的异常检测方法,其特征在于,包括:

训练对抗学习网络,得到更新的对抗学习网络,具体包括:

将正常训练样本的特征向量和模式向量输入所述对抗学习网络,利用判别器与生成器进行对抗学习,更新所述生成器和所述判别器,使所述生成器拟合所述正常训练样本的分布;

将经过插值处理的所述特征向量和所述模式向量输入所述对抗学习网络,利用编码器与所述生成器进行对抗学习,更新所述生成器和所述编码器,使所述生成器学习所述正常训练样本的潜在模式;

重复上述操作迭代训练所述对抗学习网络直至当前迭代次数达到最大迭代次数,得到所述更新的对抗学习网络;其中,所述对抗学习网络包括所述生成器、所述判别器、所述编码器;

根据训练过程中产生的重构误差构造所述更新的对抗学习网络中的异常评价函数,并将所述更新的对抗学习网络构建为异常检测模型;

通过所述异常检测模型对输入的检测数据进行异常检测,得到异常检测结果。

2.如权利要求1所述的面向多种类数据的异常检测方法,其特征在于,在所述将正常训练样本的特征向量和模式向量输入所述对抗学习网络,利用判别器与生成器进行对抗学习,更新所述生成器和所述判别器,使所述生成器拟合所述正常训练样本的分布之前,还包括:

从所述正常训练样本的特征向量集合中采集所述特征向量,并从所述正常训练样本的模式向量集合中采集与所述特征向量对应的所述模式向量。

3.如权利要求1所述的面向多种类数据的异常检测方法,其特征在于,所述将正常训练样本的特征向量和模式向量输入所述对抗学习网络,利用判别器与生成器进行对抗学习,更新所述生成器和所述判别器,使所述生成器拟合所述正常训练样本的分布,具体包括:

将所述特征向量和所述模式向量输入所述生成器,使所述生成器根据所述特征向量和所述模式向量生成第一图像;

将所述第一图像输入所述判别器,获取所述判别器对所述第一图像的异常评价;

根据所述异常评价更新所述生成器,使所述生成器拟合所述正常训练样本的分布;

将所述正常训练样本输入所述判别器,结合所述异常评价更新所述判别器。

4.如权利要求1所述的面向多种类数据的异常检测方法,其特征在于,所述将经过插值处理的所述特征向量和所述模式向量输入所述对抗学习网络,利用编码器与所述生成器进行对抗学习,更新所述生成器和所述编码器,使所述生成器学习所述正常训练样本的潜在模式,具体包括:

分别对所述特征向量和所述模式向量进行插值处理,得到插值特征向量和插值模式向量;

将所述插值特征向量和所述插值模式向量输入所述生成器,使所述生成器根据所述插值模式向量和所述插值特征向量生成第二图像和第三图像,并分别根据所述第二图像和所述第三图像计算当前的网络损失;

将所述正常训练样本输入所述编码器,使所述编码器将提取的重构特征向量和重构模式向量反向输入所述生成器,并计算当前的重构误差;

结合所述当前的网络损失和所述当前的重构误差更新所述生成器和所述编码器，使所述生成器学习所述正常训练样本的潜在模式。

5.如权利要求1所述的面向多种类数据的异常检测方法，其特征在于，所述根据训练过程中产生的重构误差构造所述更新的对抗学习网络中的异常评价函数，并将所述更新的对抗学习网络构建为异常检测模型，具体包括：

统计训练过程中产生的重构误差，将所述重构误差转换为最大似然概率；

根据所述最大似然概率构造所述更新的对抗学习网络的异常评价函数；

在构造所述异常评价函数后，将所述更新的对抗学习网络构建为所述异常检测模型。

6.一种面向多种类数据的异常检测装置，其特征在于，包括：

对抗学习网络训练模块，用于训练对抗学习网络，得到更新的对抗学习网络，具体包括：

将正常训练样本的特征向量和模式向量输入所述对抗学习网络，利用判别器与生成器进行对抗学习，更新所述生成器和所述判别器，使所述生成器拟合所述正常训练样本的分布；

将经过插值处理的所述特征向量和所述模式向量输入所述对抗学习网络，利用编码器与所述生成器进行对抗学习，更新所述生成器和所述编码器，使所述生成器学习所述正常训练样本的潜在模式；

重复上述操作迭代训练所述对抗学习网络直至当前迭代次数达到最大迭代次数，得到所述更新的对抗学习网络；其中，所述对抗学习网络包括所述生成器、所述判别器、所述编码器；

异常检测模型构建模块，用于根据训练过程中产生的重构误差构造所述更新的对抗学习网络中的异常评价函数，并将所述更新的对抗学习网络构建为异常检测模型；

数据异常检测模块，用于通过所述异常检测模型对输入的检测数据进行异常检测，得到异常检测结果。

7.如权利要求6所述的面向多种类数据的异常检测装置，其特征在于，在所述将正常训练样本的特征向量和模式向量输入所述对抗学习网络，利用判别器与生成器进行对抗学习，更新所述生成器和所述判别器，使所述生成器拟合所述正常训练样本的分布之前，还包括：

从所述正常训练样本的特征向量集合中采集所述特征向量，并从所述正常训练样本的模式向量集合中采集与所述特征向量对应的所述模式向量。

8.如权利要求6所述的面向多种类数据的异常检测装置，其特征在于，所述将正常训练样本的特征向量和模式向量输入所述对抗学习网络，利用判别器与生成器进行对抗学习，更新所述生成器和所述判别器，使所述生成器拟合所述正常训练样本的分布，具体包括：

将所述特征向量和所述模式向量输入所述生成器，使所述生成器根据所述特征向量和所述模式向量生成第一图像；

将所述第一图像输入所述判别器，获取所述判别器对所述第一图像的异常评价；

根据所述异常评价更新所述生成器，使所述生成器拟合所述正常训练样本的分布；

将所述正常训练样本输入所述判别器，结合所述异常评价更新所述判别器。

9.如权利要求6所述的面向多种类数据的异常检测装置，其特征在于，所述将经过插值

处理的所述特征向量和所述模式向量输入所述对抗学习网络，利用编码器与所述生成器进行对抗学习，更新所述生成器和所述编码器，使所述生成器学习所述正常训练样本的潜在模式，具体包括：

分别对所述特征向量和所述模式向量进行插值处理，得到插值特征向量和插值模式向量；

将所述插值特征向量和所述插值模式向量输入所述生成器，使所述生成器根据所述插值模式向量和所述插值特征向量生成第二图像和第三图像，并分别根据所述第二图像和所述第三图像计算当前的网络损失；

将所述正常训练样本输入所述编码器，使所述编码器将提取的重构特征向量和重构模式向量反向输入所述生成器，并计算当前的重构误差；

结合所述当前的网络损失和所述当前的重构误差更新所述生成器和所述编码器，使所述生成器学习所述正常训练样本的潜在模式。

10.如权利要求6所述的面向多种类数据的异常检测装置，其特征在于，所述根据训练过程中产生的重构误差构造所述更新的对抗学习网络中的异常评价函数，并将所述更新的对抗学习网络构建为异常检测模型，具体包括：

统计训练过程中产生的重构误差，将所述重构误差转换为最大似然概率；

根据所述最大似然概率构造所述更新的对抗学习网络的异常评价函数；

在构造所述异常评价函数后，将所述更新的对抗学习网络构建为所述异常检测模型。

一种面向多种类数据的异常检测方法及装置

技术领域

[0001] 本发明涉及异常检测技术领域，尤其涉及一种面向多种类数据的异常检测方法及装置。

背景技术

[0002] 近年来，随着数据量的不断增加，异常检测技术受到了广泛关注。目前提出的异常检测算法可大致分为基于近似的异常检测算法和基于模型的异常检测算法。

[0003] 基于近似的异常检测算法是根据测试样本与正常样本之间的近似程度来判断测试样本是否异常。基于近似的异常检测算法在每一次异常检测时均需要所有正常样本参与到近似程度的运算中，数据规模扩大的同时也增大了运算压力，难以应用于现实的异常检测场景。

[0004] 基于模型的异常检测算法是通过大量正常数据学习正常样本的特征和分布，并在检测阶段通过计算测试样本与正常样本的分布差异大小来判断测试样本是否异常。基于模型的异常检测算法通常将正常样本当成一类统一的模式进行学习，忽略了在现实生活中正常样本包含多种潜在模式的特性，例如在X光图片中，按性别来划分样本包含男性、女性两种模式，按体型来划分样本则包含胖瘦等模式，难以准确地捕捉到正常样本中区别于异常样本的有效模式，无法进一步提高异常检测的准确性。

发明内容

[0005] 为克服现有技术的缺陷，本发明提供一种面向多种类数据的异常检测方法及装置，能够针对多种类数据进行异常检测，进一步提高异常检测的准确性。

[0006] 为了解决上述技术问题，第一方面，本发明一实施例提供一种面向多种类数据的异常检测方法，包括：

[0007] 训练对抗学习网络，得到更新的对抗学习网络，具体包括：

[0008] 将正常训练样本的特征向量和模式向量输入所述对抗学习网络，利用判别器与生成器进行对抗学习，更新所述生成器和所述判别器，使所述生成器拟合所述正常训练样本的分布；

[0009] 将经过插值处理的所述特征向量和所述模式向量输入所述对抗学习网络，利用编码器与所述生成器进行对抗学习，更新所述生成器和所述编码器，使所述生成器学习所述正常训练样本的潜在模式；

[0010] 重复上述操作迭代训练所述对抗学习网络直至当前迭代次数达到最大迭代次数，得到所述更新的对抗学习网络；其中，所述对抗学习网络包括所述生成器、所述判别器、所述编码器；

[0011] 根据训练过程中产生的重构误差构造所述更新的对抗学习网络中的异常评价函数，并将所述更新的对抗学习网络构建为异常检测模型；

[0012] 通过所述异常检测模型对输入的检测数据进行异常检测，得到异常检测结果。

[0013] 进一步地，在所述将正常训练样本的特征向量和模式向量输入所述对抗学习网络，利用判别器与生成器进行对抗学习，更新所述生成器和所述判别器，使所述生成器拟合所述正常训练样本的分布之前，还包括：

[0014] 从所述正常训练样本的特征向量集合中采集所述特征向量，并从所述正常训练样本的模式向量集合中采集与所述特征向量对应的所述模式向量。

[0015] 进一步地，所述将正常训练样本的特征向量和模式向量输入所述对抗学习网络，利用判别器与生成器进行对抗学习，更新所述生成器和所述判别器，使所述生成器拟合所述正常训练样本的分布，具体包括：

[0016] 将所述特征向量和所述模式向量输入所述生成器，使所述生成器根据所述特征向量和所述模式向量生成第一图像；

[0017] 将所述第一图像输入所述判别器，获取所述判别器对所述第一图像的异常评价；

[0018] 根据所述异常评价更新所述生成器，使所述生成器拟合所述正常训练样本的分布；

[0019] 将所述正常训练样本输入所述判别器，结合所述异常评价更新所述判别器。

[0020] 进一步地，所述将经过插值处理的所述特征向量和所述模式向量输入所述对抗学习网络，利用编码器与所述生成器进行对抗学习，更新所述生成器和所述编码器，使所述生成器学习所述正常训练样本的潜在模式，具体包括：

[0021] 分别对所述特征向量和所述模式向量进行插值处理，得到插值特征向量和插值模式向量；

[0022] 将所述插值特征向量和所述插值模式向量输入所述生成器，使所述生成器根据所述插值模式向量和所述插值特征向量生成第二图像和第三图像，并分别根据所述第二图像和所述第三图像计算当前的网络损失；

[0023] 将所述正常训练样本输入所述编码器，使所述编码器将提取的重构特征向量和重构模式向量反向输入所述生成器，并计算当前的重构误差；

[0024] 结合所述当前的网络损失和所述当前的重构误差更新所述生成器和所述编码器，使所述生成器学习所述正常训练样本的潜在模式。

[0025] 进一步地，所述根据训练过程中产生的重构误差构造所述更新的对抗学习网络中的异常评价函数，并将所述更新的对抗学习网络构建为异常检测模型，具体包括：

[0026] 统计训练过程中产生的重构误差，将所述重构误差转换为最大似然概率；

[0027] 根据所述最大似然概率构造所述更新的对抗学习网络的异常评价函数；

[0028] 在构造所述异常评价函数后，将所述更新的对抗学习网络构建为所述异常检测模型。

[0029] 第二方面，本发明一实施例提供一种面向多种类数据的异常检测装置，包括：

[0030] 对抗学习网络训练模块，用于训练对抗学习网络，得到更新的对抗学习网络，具体包括：

[0031] 将正常训练样本的特征向量和模式向量输入所述对抗学习网络，利用判别器与生成器进行对抗学习，更新所述生成器和所述判别器，使所述生成器拟合所述正常训练样本的分布；

[0032] 将经过插值处理的所述特征向量和所述模式向量输入所述对抗学习网络，利用编

码器与所述生成器进行对抗学习，更新所述生成器和所述编码器，使所述生成器学习所述正常训练样本的潜在模式；

[0033] 重复上述操作迭代训练所述对抗学习网络直至当前迭代次数达到最大迭代次数，得到所述更新的对抗学习网络；其中，所述对抗学习网络包括所述生成器、所述判别器、所述编码器；

[0034] 异常检测模型构建模块，用于根据训练过程中产生的重构误差构造所述更新的对抗学习网络中的异常评价函数，并将所述更新的对抗学习网络构建为异常检测模型；

[0035] 数据异常检测模块，用于通过所述异常检测模型对输入的检测数据进行异常检测，得到异常检测结果。

[0036] 进一步地，在所述将正常训练样本的特征向量和模式向量输入所述对抗学习网络，利用判别器与生成器进行对抗学习，更新所述生成器和所述判别器，使所述生成器拟合所述正常训练样本的分布之前，还包括：

[0037] 从所述正常训练样本的特征向量集合中采集所述特征向量，并从所述正常训练样本的模式向量集合中采集与所述特征向量对应的所述模式向量。

[0038] 进一步地，所述将正常训练样本的特征向量和模式向量输入所述对抗学习网络，利用判别器与生成器进行对抗学习，更新所述生成器和所述判别器，使所述生成器拟合所述正常训练样本的分布，具体包括：

[0039] 将所述特征向量和所述模式向量输入所述生成器，使所述生成器根据所述特征向量和所述模式向量生成第一图像；

[0040] 将所述第一图像输入所述判别器，获取所述判别器对所述第一图像的异常评价；

[0041] 根据所述异常评价更新所述生成器，使所述生成器拟合所述正常训练样本的分布；

[0042] 将所述正常训练样本输入所述判别器，结合所述异常评价更新所述判别器。

[0043] 进一步地，所述将经过插值处理的所述特征向量和所述模式向量输入所述对抗学习网络，利用编码器与所述生成器进行对抗学习，更新所述生成器和所述编码器，使所述生成器学习所述正常训练样本的潜在模式，具体包括：

[0044] 分别对所述特征向量和所述模式向量进行插值处理，得到插值特征向量和插值模式向量；

[0045] 将所述插值特征向量和所述插值模式向量输入所述生成器，使所述生成器根据所述插值模式向量和所述插值特征向量生成第二图像和第三图像，并分别根据所述第二图像和所述第三图像计算当前的网络损失；

[0046] 将所述正常训练样本输入所述编码器，使所述编码器将提取的重构特征向量和重构模式向量反向输入所述生成器，并计算当前的重构误差；

[0047] 结合所述当前的网络损失和所述当前的重构误差更新所述生成器和所述编码器，使所述生成器学习所述正常训练样本的潜在模式。

[0048] 进一步地，所述根据训练过程中产生的重构误差构造所述更新的对抗学习网络中的异常评价函数，并将所述更新的对抗学习网络构建为异常检测模型，具体包括：

[0049] 统计训练过程中产生的重构误差，将所述重构误差转换为最大似然概率；

[0050] 根据所述最大似然概率构造所述更新的对抗学习网络的异常评价函数；

[0051] 在构造所述异常评价函数后，将所述更新的对抗学习网络构建为所述异常检测模型。

[0052] 本发明的实施例，具有如下有益效果：

[0053] 通过训练对抗学习网络，使对抗学习网络中的生成器拟合正常训练样本的分布以及学习正常训练样本的潜在模式，得到更新的对抗学习网络，根据训练过程中产生的重构误差构造更新的对抗学习网络中的异常评价函数，并将更新的对抗学习网络构建为异常检测模型，以利用异常检测模型对输入的检测数据进行异常检测，得到异常检测结果，从而能够对多种类数据进行异常检测，提高异常检测的准确性。

附图说明

[0054] 图1为本发明第一实施例中的一种面向多种类数据的异常检测方法的流程示意图；

[0055] 图2为本发明第一实施例中的对抗学习网络的结构示意图；

[0056] 图3为本发明第一实施例中的对抗学习网络的训练算法流程图；

[0057] 图4为本发明第二实施例中的一种面向多种类数据的异常检测装置的结构示意图。

具体实施方式

[0058] 下面将结合本发明中的附图，对本发明中的技术方案进行清楚、完整地描述，显然，所描述的实施例仅仅是本发明一部分实施例，而不是全部的实施例。基于本发明中的实施例，本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其他实施例，都属于本发明保护的范围。

[0059] 需要说明的是，文中的步骤编号，仅为了方便具体实施例的解释，不作为限定步骤执行先后顺序的作用。

[0060] 第一实施例：

[0061] 如图1所示，第一实施例提供一种面向多种类数据的异常检测方法，包括步骤S1～S3：

[0062] S1、训练对抗学习网络，得到更新的对抗学习网络，具体包括：

[0063] 将正常训练样本的特征向量和模式向量输入对抗学习网络，利用判别器与生成器进行对抗学习，更新生成器和判别器，使生成器拟合正常训练样本的分布；

[0064] 将经过插值处理的特征向量和模式向量输入对抗学习网络，利用编码器与生成器进行对抗学习，更新生成器和编码器，使生成器学习正常训练样本的潜在模式；

[0065] 重复上述操作迭代训练对抗学习网络直至当前迭代次数达到最大迭代次数，得到更新的对抗学习网络；其中，对抗学习网络包括生成器、判别器、编码器；

[0066] S2、根据训练过程中产生的重构误差构造更新的对抗学习网络中的异常评价函数，并将更新的对抗学习网络构建为异常检测模型；

[0067] S3、通过异常检测模型对输入的检测数据进行异常检测，得到异常检测结果。

[0068] 需要说明的是，正常训练样本是从正常数据中筛选获得的。

[0069] 如图2所示，假设正常数据的分布是由多种模式组成，可以将正常数据表征为特征

向量z_n和模式向量z_c。其中，特征向量z_n是一个d维的向量，用来指示正常数据的特征表示，记作$z_n \in R_d$；模式向量z_c是一个c维的0-1向量(one-hot 向量)，用来指示数据的模式类别，记作$z_c \in E_t$，后文均使用E_t来表示0-1编码集。假设正常数据中，数据之所以可以被划分为同一种模式，是因为这些数据具有相似的特征，从而可以学到共同的特征表示，因此每一个正常数据的模式均可以使用一个高斯分布表征它的特征，即$z_{ni} \sim N(0,\sigma^2)$，i＝1,...,d。为了书写的方便，本实施例将特征向量z_n记作$z_n \in N(0,\sigma^2)$，经过插值处理的特征向量z_n和模式向量z_c分别表示为插值特征向量$\mathbf{z}_n^{int}$和插值模式向量$\mathbf{z}_c^{int}$。对抗学习网络包括生成器(Generator)、判别器(Discriminator)和编码器(Encoder)。生成器的功能是根据输入的特征向量和模式向量生成对应图像，即$G(\cdot,\cdot):(R^c,R^d)\rightarrow R^n$，其中，$R^c$表示c维向量的集合，生成器的主要目标是拟合正常训练样本的分布，并且学习到正常训练样本的潜在模式。而编码器反向地学习提取输入图像的模式及表征，即$E:R^n\rightarrow(R^c,R^d)$，因此可以通过学习到的编码器，提取输入图像的模式信息和特征表示信息，然后输入到生成器进行对该输入图像的重构，同时编码器对于模式信息的提取也可以指导生成器对潜在模式的学习。判别器主要是通过与生成器的对抗学习，指导生成器拟合正常数据的分布。

[0070] 对抗学习网络的训练过程主要分为两个阶段来迭代。第一阶段的主要目的是使生成器拟合正常训练样本的分布以及学习正常训练样本的潜在模式，生成器通过与判别器对抗学习，拟合正常训练样本的分布，同时通过编码器学习正常训练样本的潜在模式。第二阶段的主要目的是评估第一步中所学习到的潜在模式的质量，通过验证当前学习到的潜在模式是否符合异常检测有利的模式的特性，将有利于异常检测的模式信息学习出来。整个训练过程会不断重复这两个阶段，指导生成器能成功拟合正常训练样本的分布，并且学习到的模式收敛至对异常检测有帮助的模式上。

[0071] 在优选的实施例当中，在将正常训练样本的特征向量和模式向量输入对抗学习网络，利用判别器与生成器进行对抗学习，更新生成器和判别器，使生成器拟合正常训练样本的分布之前，还包括：从正常训练样本的特征向量集合中采集特征向量，并从正常训练样本的模式向量集合中采集与特征向量对应的模式向量。

[0072] 在优选的实施例当中，将正常训练样本的特征向量和模式向量输入对抗学习网络，利用判别器与生成器进行对抗学习，更新生成器和判别器，使生成器拟合正常训练样本的分布，具体包括：将特征向量和模式向量输入生成器，使生成器根据特征向量和模式向量生成第一图像；将第一图像输入判别器，获取判别器对第一图像的异常评价；根据异常评价更新生成器，使生成器拟合正常训练样本的分布；将正常训练样本输入判别器，结合异常评价更新判别器。

[0073] 在优选的实施例当中，将经过插值处理的特征向量和模式向量输入对抗学习网络，利用编码器与生成器进行对抗学习，更新生成器和编码器，使生成器学习正常训练样本的潜在模式，具体包括：分别对特征向量和模式向量进行插值处理，得到插值特征向量和插值模式向量；将插值特征向量和插值模式向量输入生成器，使生成器根据插值模式向量和插值特征向量生成第二图像和第三图像，并分别根据第二图像和第三图像计算当前的网络损失；将正常训练样本输入编码器，使编码器将提取的重构特征向量和重构模式向量反向输入生成器，并计算当前的重构误差；结合当前的网络损失和当前的重构误差更新生成器和编码器，使生成器学习正常训练样本的潜在模式。

[0074] 作为示例性地，对抗学习网络的训练算法流程图如图3所示。在训练对抗学习网络之前，需要对生成器G、编码器E和判别器D进行初始化，iter是当前迭代的步数，nepochs为设定的最大迭代步数。

[0075] 首先从$N(0,\sigma^2)$中采样一个批次的特征向量z_n，同时从均匀分布U[1,c]中采样出这个批次需要生成的模式类别，然后转化成one-hot向量，得到模式向量z_c。

[0076] 然后将获得的特征向量z_n和模式向量z_c输入生成器，使生成器根据特征向量z_n和模式向量z_c生成一个批次的假图片X_{gen}(即第一图像)。将这批假图片输入判别器，获取判别器的评价，然后更新生成器使生成器拟合正常训练样本的分布。将真实的正常训练样本输入判别器，结合刚才的假图像评价，更新判别器。这个过程是生成器和判别器之间的对抗学习。

[0077] 接下来进行插值处理和对第一阶段学习到的模式进行评价。通过对特征向量z_n进行插值处理，获得相同模式内的插值特征向量$z_n^{int}=\sum\lambda_i z_n^i$和插值模式向量$z_c^{int}=z_c$，通过对模式向量$z_c$进行插值处理，获得不同模式间插值特征向量$z_n^{int}=z_n$和插值模式向量$z_c^{int}$。将这些向量输入生成器，使生成器生成对应的相同模式内插值图像X_{same}(即第二图像)和不同模式间插值图像X_{dif}(即第三图像)。根据不同插值处理的特性运算对应的网络损失L_{cycle}和$L_{inter\ polation}$。将真实的正常训练样本输入编码器，使编码器将从中提取的重构特征向量和重构模式向量输入生成器，求得当前的重构误差。综合这些网络损失和重构误差，更新生成器和编码器，惩罚第一阶段学习到的无效的模式信息。

[0078] 重复这些步骤直到对抗学习网络模型收敛或者当前的迭代次数到达最大迭代次数。

[0079] 图3中，公式(1)如下所示：

[0080] $$L_{adv}=E_{z_c\in e_t,z_n\sim N(0,\sigma^2)}\log(1-D(G(z_c,z_n)))+E_{X\sim P_n(x)}\log(D(X)) \quad (1)；$$

[0081] 公式(1)中，L_{adv}为当前对抗生成网络总体的损失函数；e_t为所有类别样本的集合；$G(z_c,z_n)$为生成器针对z_c，z_n的输出图像，$D(G(z_c,z_n))$为判别器对生成器的输出图像的评价；$P_n(x)=G^*(z_c,z_n)$，G^*为当前最优的生成器；$D(X)$为判别器对当前最优生成器的输出图像的评价；

[0082] 公式(2)如下所示：

[0083] $$L_{cycle}=E_{z_c\in e_t,z_n\sim N(0,\sigma^2)}H(\widehat{z_c},z_c)+\|\widehat{z_n}-z_n\| \quad (2)；$$

[0084] 公式(2)中，e_t为所有类别样本的集合；H是交叉熵，$(\widehat{z_c},\widehat{z_n})=E(G(z_c,z_n))$，$\widehat{z_c}$和$\widehat{z_n}$分别是编码器输出的重构模式向量和重构特征向量；

[0085] 公式(3)如下所示：

[0086] $$L_{interpolation}=E_{z_c\in e_t z_n}\sum-\frac{1}{c}log\widehat{z_c},\quad where\ \widehat{z_c},\ \widehat{z_n}=E(G(z_c,z_n)) \quad (3)；$$

[0087] 公式(3)中，$E_{z_c\notin e_t,z_n}$为不属于学习的样本(z_c,z_n)；c为类别数量；

[0088] 假设z_c是100维的one-hot向量，能够表示100种不同类别，但实际上只需要学习其

中的某10种类别。在进行插值处理随机生成z_c的过程中，很有可能会生成这10种类别之外的z_c，此时需要通过这个损失函数来惩罚生成这些类别之外的向量，因为这些向量是无意义的。

[0089] 公式(4)如下所示：

[0090] $L=\alpha L_{intepolation}+\beta L_{cycle}+\gamma L_{recon}$ (4)；

[0091] 公式(4)中，α，β，γ均为预设的超参数。

[0092] 在优选的实施例当中，根据训练过程中产生的重构误差构造更新的对抗学习网络中的异常评价函数，并将更新的对抗学习网络构建为异常检测模型，具体包括：统计训练过程中产生的重构误差，将重构误差转换为最大似然概率；根据最大似然概率构造更新的对抗学习网络的异常评价函数；在构造异常评价函数后，将更新的对抗学习网络构建为异常检测模型。

[0093] 异常检测模型按功能主要划分为一个生成网络p(x;θ)和模式分类网络 q(x;φ)，可通过一个异常评价函数对样本异常程度进行打分，异常评价函数的运算公式如公式(5)所示：

[0094] $$A(x)=\left(1-logp(x;\theta)\right)*\frac{1}{n_c}-logp(x;\theta)logq(x;\phi) \quad (5)；$$

[0095] 公式(5)中，n_c为划分的模式数目，

$$0\le logp\left(p(x;\theta)\right)\le 1,\frac{1}{n_c}\le log\left(q(x;\phi)\right)\le 1$$。

[0096] 异常评价函数的取值范围为-1≤A(x)≤-1，其中数值越大，表明样本的异常程度越高。这个异常评价函数实际上将数据分成了正常数据、远离正常分布的异常数据、靠近正常分布的异常数据三大类。对于正常数据有logp(x;θ)＝1，并且由于正常数据可以被大概率划分为模式中的任一种模式，有logq(x;φ)＝1，因此有A(x)＝-1。而对于远离正常分布的异常数据而言，$logp(x;\theta)=0, logq(x;\theta)=\frac{1}{n_c}$，因此$A(x)=-\frac{1}{n_c}$。同理，对于与正常分布距离较近的异常数据，由于生成器失效无法将其区别出来，有log(p(x;φ))＝1。此时A(x)主要由第二项决定，由于分类器对于此类异常数据可以成功将其划分为不属于任何一种模式而区别出来，有A(x)＝-1。异常判断的依据在生成网络和模式分类网络上的权重主要由数据与正常分布的距离来决定。由于logp(x;θ)能一定程度上反映当前数据与正常分布的距离，因此当距离越近时，logp(x;θ)越接近1，则A(x)中第二项的权重越大，因此判断的依据越依赖于模式分类器。相反，当数据于正常分布明显较远，差异较大时，A(x)主要依赖第一项，克服了模式分类器失效而输出的高置信度信号。

[0097] 在异常检测模型的具体实现中，由于logp(x;θ)与重构误差的大小成反比，输入样本的重构误差越小，说明该样本越有可能是正常样本，越接近正常样本的分布。为了建立一个概率上的度量，通过对所有正常训练样本的重构误差进行统计，从而获得一个先验的从重构误差到最大似然概率间的转换，如公式(6) 所示：

[0098] $log(p(x;\theta))\approx P_N(||X-G(E(X))||)$ (6)；

[0099] 公式(6)中，PN(•)是服从均值为μ，方差为σ的正态分布。μ，σ是通过所有的正常训练样本的重构误差进行统计求得。而对于模式分类网络的概率，主要通过分类器输出的属

于各类概率的最大值来实现，如公式(7)所示：

[0100] $$log\left(q(x;\phi)\right) \approx -argmax\left(P_{x\in c_i}(x)\right) \quad (7)；$$

[0101] 综上，异常检测模型最终构造的异常评价函数如公式(8)所示：

[0102] $$A(x) = -\left(1 - P_N\left(\left\|X - G(E(X))\right\|\right)\right) * \frac{1}{n_c} - P_N\left(\left\|X - G(E(X))\right\|\right) \arg\max P_{x\in c_i}(X) \quad (8)。$$

[0103] 第二实施例：

[0104] 如图4所示，第二实施例提供一种面向多种类数据的异常检测装置，包括：对抗学习网络训练模块21，用于训练对抗学习网络，得到更新的对抗学习网络，具体包括：将正常训练样本的特征向量和模式向量输入对抗学习网络，利用判别器与生成器进行对抗学习，更新生成器和判别器，使生成器拟合正常训练样本的分布；将经过插值处理的特征向量和模式向量输入对抗学习网络，利用编码器与生成器进行对抗学习，更新生成器和编码器，使生成器学习正常训练样本的潜在模式；重复上述操作迭代训练对抗学习网络直至当前迭代次数达到最大迭代次数，得到更新的对抗学习网络；其中，对抗学习网络包括生成器、判别器、编码器；异常检测模型构建模块22，用于根据训练过程中产生的重构误差构造更新的对抗学习网络中的异常评价函数，并将更新的对抗学习网络构建为异常检测模型；数据异常检测模块23，用于通过异常检测模型对输入的检测数据进行异常检测，得到异常检测结果。

[0105] 需要说明的是，正常训练样本是从正常数据中筛选获得的。

[0106] 假设正常数据的分布是由多种模式组成，可以将正常数据表征为特征向量 z_n和模式向量z_c。其中，特征向量z_n是一个d维的向量，用来指示正常数据的特征表示，记作$z_n \in R_d$；模式向量z_c是一个c维的0-1向量(one-hot向量)，用来指示数据的模式类别，记作$z_c \in E_t$，后文均使用E_t来表示0-1编码集。假设正常数据中，数据之所以可以被划分为同一种模式，是因为这些数据具有相似的特征，从而可以学到共同的特征表示，因此每一个正常数据的模式均可以使用一个高斯分布表征它的特征，即$z_{ni} \sim N(0,\sigma^2)$，i＝1，...，d。为了书写的方便，本实施例将特征向量z_n记作$z_n \in N(0,\sigma^2)$，经过插值处理的特征向量z_n和模式向量z_c分别表示为插值特征向量z_n^{int}和插值模式向量z_c^{int}。对抗学习网络包括生成器 (Generator)、判别器(Discriminator)和编码器(Encoder)。生成器的功能是根据输入的特征向量和模式向量生成对应图像，即$G(\cdot,\cdot):(R^c,R^d)\rightarrow R^n$，其中，$R^c$表示c维向量的集合，生成器的主要目标是拟合正常训练样本的分布，并且学习到正常训练样本的潜在模式。而编码器反向地学习提取输入图像的模式及表征，即$E:R^n\rightarrow(R^c,R^d)$，因此可以通过学习到的编码器，提取输入图像的模式信息和特征表示信息，然后输入到生成器进行对该输入图像的重构，同时编码器对于模式信息的提取也可以指导生成器对潜在模式的学习。判别器主要是通过与生成器的对抗学习，指导生成器拟合正常数据的分布。

[0107] 对抗学习网络训练模块21对于对抗学习网络的训练过程主要分为两个阶段来迭代。第一阶段的主要目的是使生成器拟合正常训练样本的分布以及学习正常训练样本的潜在模式，生成器通过与判别器对抗学习，拟合正常训练样本的分布，同时通过编码器学习正常训练样本的潜在模式。第二阶段的主要目的是评估第一步中所学习到的潜在模式的质量，通过验证当前学习到的潜在模式是否符合异常检测有利的模式的特性，将有利于异常检测的模式信息学习出来。整个训练过程会不断重复这两个阶段，指导生成器能成功拟合

正常训练样本的分布，并且学习到的模式收敛至对异常检测有帮助的模式上。

[0108] 在优选的实施例当中，在将正常训练样本的特征向量和模式向量输入对抗学习网络，利用判别器与生成器进行对抗学习，更新生成器和判别器，使生成器拟合正常训练样本的分布之前，还包括：从正常训练样本的特征向量集合中采集特征向量，并从正常训练样本的模式向量集合中采集与特征向量对应的模式向量。

[0109] 在优选的实施例当中，将正常训练样本的特征向量和模式向量输入对抗学习网络，利用判别器与生成器进行对抗学习，更新生成器和判别器，使生成器拟合正常训练样本的分布，具体包括：将特征向量和模式向量输入生成器，使生成器根据特征向量和模式向量生成第一图像；将第一图像输入判别器，获取判别器对第一图像的异常评价；根据异常评价更新生成器，使生成器拟合正常训练样本的分布；将正常训练样本输入判别器，结合异常评价更新判别器。

[0110] 在优选的实施例当中，将经过插值处理的特征向量和模式向量输入对抗学习网络，利用编码器与生成器进行对抗学习，更新生成器和编码器，使生成器学习正常训练样本的潜在模式，具体包括：分别对特征向量和模式向量进行插值处理，得到插值特征向量和插值模式向量；将插值特征向量和插值模式向量输入生成器，使生成器根据插值模式向量和插值特征向量生成第二图像和第三图像，并分别根据第二图像和第三图像计算当前的网络损失；将正常训练样本输入编码器，使编码器将提取的重构特征向量和重构模式向量反向输入生成器，并计算当前的重构误差；结合当前的网络损失和当前的重构误差更新生成器和编码器，使生成器学习正常训练样本的潜在模式。

[0111] 作为示例性地，对抗学习网络训练模块21在训练对抗学习网络之前，需要对生成器G、编码器E和判别器D进行初始化。

[0112] 首先从$N(0,\sigma^2)$中采样一个批次的特征向量z_n，同时从均匀分布$U[1,c]$中采样出这个批次需要生成的模式类别，然后转化成one-hot向量，得到模式向量z_c。

[0113] 然后将获得的特征向量z_n和模式向量z_c输入生成器，使生成器根据特征向量z_n和模式向量z_c生成一个批次的假图片X_{gen}(即第一图像)。将这批假图片输入判别器，获取判别器的评价，然后更新生成器使生成器拟合正常训练样本的分布。将真实的正常训练样本输入判别器，结合刚才的假图像评价，更新判别器。这个过程是生成器和判别器之间的对抗学习。

[0114] 接下来进行插值处理和对第一阶段学习到的模式进行评价。通过对特征向量z_n进行插值处理，获得相同模式内的插值特征向量$z_n^{int}=\sum\lambda_i z_n^i$和插值模式向量$z_c^{int}=z_c$，通过对模式向量$z_c$进行插值处理，获得不同模式间插值特征向量$z_n^{int}=z_n$和插值模式向量$z_c^{int}$。将这些向量输入生成器，使生成器生成对应的相同模式内插值图像X_{same}(即第二图像)和不同模式间插值图像X_{dif}(即第三图像)。根据不同插值处理的特性运算对应的网络损失L_{cycle}和$L_{inter\ polation}$。将真实的正常训练样本输入编码器，使编码器将从中提取的重构特征向量和重构模式向量输入生成器，求得当前的重构误差。综合这些网络损失和重构误差，更新生成器和编码器，惩罚第一阶段学习到的无效的模式信息。

[0115] 重复这些步骤直到对抗学习网络模型收敛或者当前的迭代次数到达最大迭代次数。

[0116] 在优选的实施例当中，根据训练过程中产生的重构误差构造更新的对抗学习网络中的异常评价函数，并将更新的对抗学习网络构建为异常检测模型，具体包括：统计训练过程中产生的重构误差，将重构误差转换为最大似然概率；根据最大似然概率构造更新的对抗学习网络的异常评价函数；在构造异常评价函数后，将更新的对抗学习网络构建为异常检测模型。

[0117] 异常检测模型按功能主要划分为一个生成网络p(x;θ)和模式分类网络 q(x;φ)，可通过一个异常评价函数对样本异常程度进行打分，异常评价函数的运算公式如公式(9)所示：

[0118] $$A(x)=\left(1-logp(x;\theta)\right)*\frac{1}{n_c}-logp(x;\theta)logq(x;\phi) \quad (9)；$$

[0119] 公式(9)中，n_c为划分的模式数目，

$$0\le logp\left(p(x;\theta)\right)\le 1,\frac{1}{n_c}\le log\left(q(x;\phi)\right)\le 1$$ 。

[0120] 异常评价函数的取值范围为-1≤A(x)≤-1，其中数值越大，表明样本的异常程度越高。这个异常评价函数实际上将数据分成了正常数据、远离正常分布的异常数据、靠近正常分布的异常数据三大类。对于正常数据有logp(x;θ)＝1，并且由于正常数据可以被大概率划分为模式中的任一种模式，有logq(x;φ)＝1，因此有A(x)＝-1。而对于远离正常分布的异常数据而言，$logp(x;\theta)=0,logq(x;\theta)=\frac{1}{n_c}$，因此$A(x)=-\frac{1}{n_c}$。同理，对于与正常分布距离较近的异常数据，由于生成器失效无法将其区别出来，有log(p(x;φ))＝1。此时A(x)主要由第二项决定，由于分类器对于此类异常数据可以成功将其划分为不属于任何一种模式而区别出来，有A(x)＝-1。异常判断的依据在生成网络和模式分类网络上的权重主要由数据与正常分布的距离来决定。由于logp(x;θ)能一定程度上反映当前数据与正常分布的距离，因此当距离越近时，logp(x;θ)越接近1，则A(x)中第二项的权重越大，因此判断的依据越依赖于模式分类器。相反，当数据于正常分布明显较远，差异较大时，A(x)主要依赖第一项，克服了模式分类器失效而输出的高置信度信号。

[0121] 在异常检测模型的具体实现中，由于logp(x;θ)与重构误差的大小成反比，输入样本的重构误差越小，说明该样本越有可能是正常样本，越接近正常样本的分布。为了建立一个概率上的度量，通过对所有正常训练样本的重构误差进行统计，从而获得一个先验的从重构误差到最大似然概率间的转换，如公式(10) 所示：

[0122] $log(p(x;\theta))\approx P_N(||X-G(E(X))||)$ (10)；

[0123] 公式(6)中，PN(•)是服从均值为μ，方差为σ的正态分布。μ，σ是通过所有的正常训练样本的重构误差进行统计求得。而对于模式分类网络的概率，主要通过分类器输出的属于各类概率的最大值来实现，如公式(11)所示：

[0124] $$log\left(q(x;\phi)\right)\approx -argmax\left(P_{x\in c_i}(x)\right) \quad (11)；$$

[0125] 综上，异常检测模型最终构造的异常评价函数如公式(12)所示：

[0126] $$A(x)=-\left(1-P_N\left(\left\|X-G(E(X))\right\|\right)\right)*\frac{1}{n_c}-P_N\left(\left\|X-G(E(X))\right\|\right)\arg\max P_{x\in c_i}(X) \quad (12)。$$

[0127] 综上所述，实施本发明的实施例，具有如下有益效果：

[0128] 通过训练对抗学习网络，使对抗学习网络中的生成器拟合正常训练样本的分布以及学习正常训练样本的潜在模式，得到更新的对抗学习网络，根据训练过程中产生的重构误差构造更新的对抗学习网络中的异常评价函数，并将更新的对抗学习网络构建为异常检测模型，以利用异常检测模型对输入的检测数据进行异常检测，得到异常检测结果，从而能够对多种类数据进行异常检测，提高异常检测的准确性。

[0129] 以上所述是本发明的优选实施方式，应当指出，对于本技术领域的普通技术人员来说，在不脱离本发明原理的前提下，还可以做出若干改进和润饰，这些改进和润饰也视为本发明的保护范围。

[0130] 本领域普通技术人员可以理解实现上述实施例中的全部或部分流程，是可以通过计算机程序来指令相关的硬件来完成，所述的程序可存储于一计算机可读取存储介质中，该程序在执行时，可包括如上述各实施例的流程。其中，所述的存储介质可为磁碟、光盘、只读存储记忆体(Read-Only Memory，ROM) 或随机存储记忆体(Random Access Memory，RAM)等。

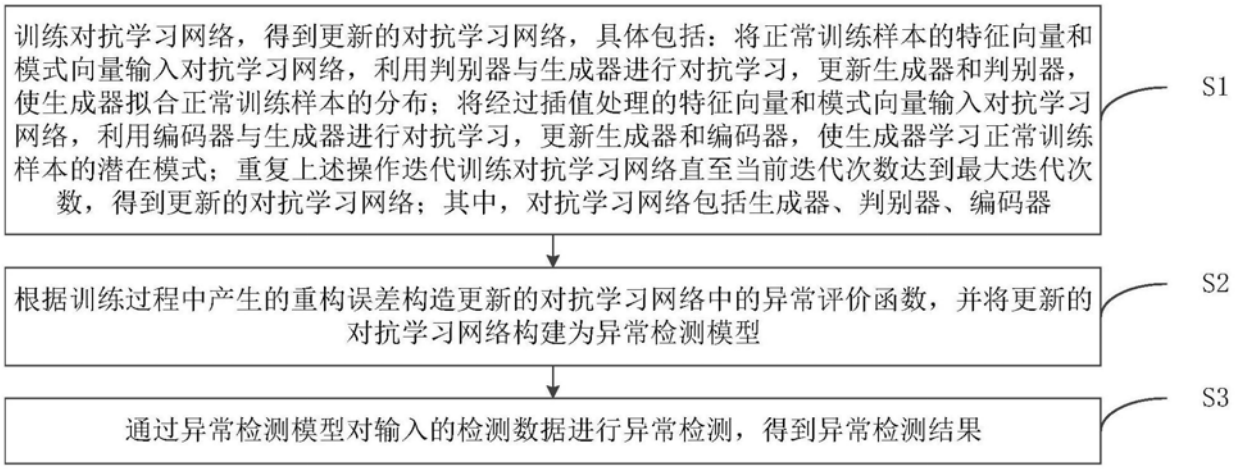

训练对抗学习网络，得到更新的对抗学习网络，具体包括：将正常训练样本的特征向量和模式向量输入对抗学习网络，利用判别器与生成器进行对抗学习，更新生成器和判别器，使生成器拟合正常训练样本的分布；将经过插值处理的特征向量和模式向量输入对抗学习网络，利用编码器与生成器进行对抗学习，更新生成器和编码器，使生成器学习正常训练样本的潜在模式；重复上述操作迭代训练对抗学习网络直至当前迭代次数达到最大迭代次数，得到更新的对抗学习网络；其中，对抗学习网络包括生成器、判别器、编码器
S1
根据训练过程中产生的重构误差构造更新的对抗学习网络中的异常评价函数，并将更新的对抗学习网络构建为异常检测模型
S2
通过异常检测模型对输入的检测数据进行异常检测，得到异常检测结果
S3

图1

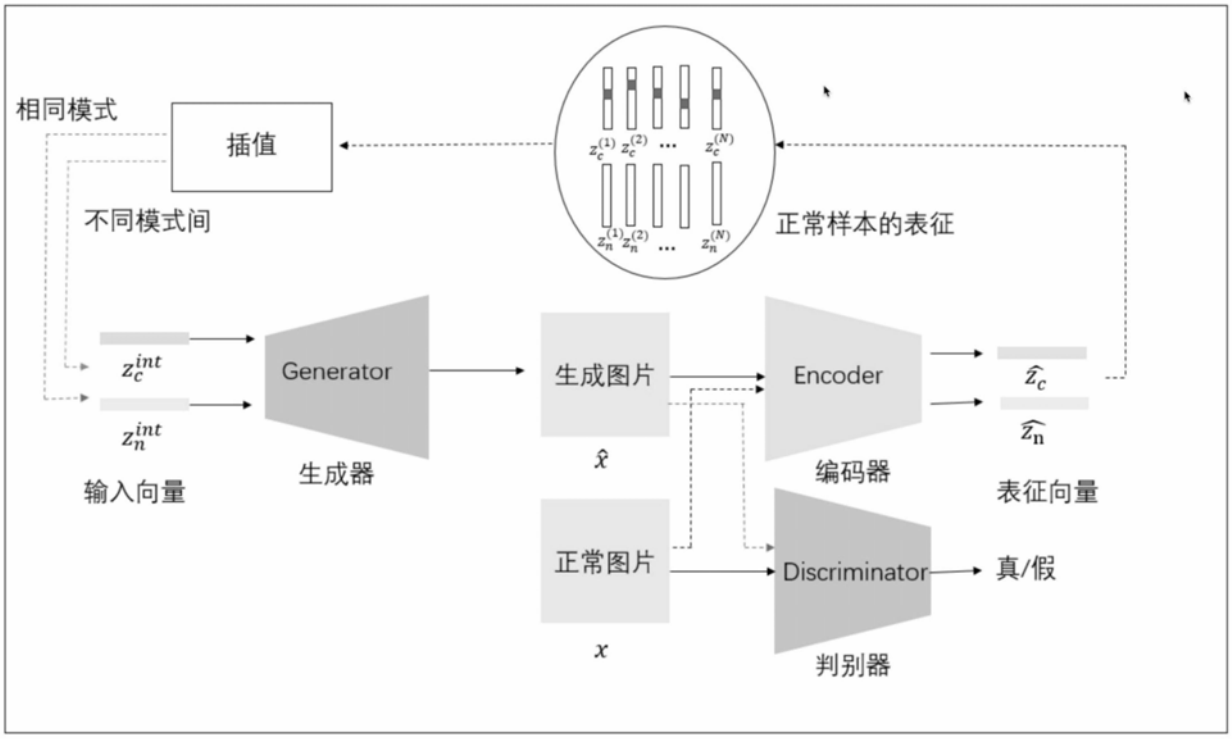

相同模式
插值
不同模式间
正常样本的表征
Generator
生成器
输入向量
生成图片
Encoder
编码器
表征向量
正常图片
Discriminator
真/假
判别器

图2

Input: X,G,E,D,iter,nepochs

Output: G,E,D

初始化 $G,E,D,iter = 1$,nepochs,σ,c;

repeat

  采样 z_c 和 z_n,$z_c \sim N(0,\sigma^2)$, $z_n \sim U[1,c]$;

  生成插值图像 $X_{gen} = G(z_c, z_n)$;

  计算 $D(X_{gen})$, 然后根据公式(1) 更新 G;

  计算 $D(X)$, 然后根据公式(1) 更新 D;

  执行插值操作 $z_c^{int} = \sum \lambda_i z_c^{(i)}$, $z_n^{int} = \sum \lambda_i z_n^{(i)}$;

  生成相同模式及不同模式间插值图像:

  $X_{same} = G(z_n^{int}, z_c), X_{dif} = G(z_n, z_c^{int})$;

  对 X_{same} 利用公式 (2) 求 $\mathcal{L}_{cycle}$;

  求重构误差,$\mathcal{L}_{recon} = \|X - G(E(X))\|$;

  对 X_{dif} 根据公式 (3) 求插值间误差 $\mathcal{L}_{interpolation}$;

  根据公式(4) 更新 G, E;

until *iter<nepochs*;

图3

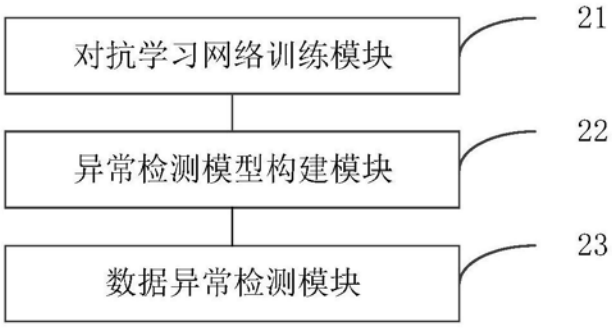


图4