



# 基于扩散常微分方程的医学图像异常检测

胡显耀, 靳聪明

(浙江理工大学理学院, 杭州 310018)

**摘要:** 为了识别与正常生理状态显著偏离的异常医学图像, 基于异常图像相关特征往往分布于特征空间低密度区域的假设, 提出了一种神经网络辅助的基于扩散常微分方程(Ordinary differential equations, ODE)的医学图像异常检测方法。该方法首先利用扩散 ODE 估计图像特征的似然值; 然后构建神经网络, 拟合图像特征在不同时刻由扩散 ODE 估计的似然值; 最后通过扩散 ODE 估计的似然值和神经网络估计的似然值的加权平均得到异常分数, 异常分数较大的图像被认定为异常图像。此外, 为了确定异常图像的异常区域, 提出了一种基于图像重构的异常定位方法, 通过计算重构误差来定位异常区域。在 BraTS2021 脑肿瘤数据集和 X 射线胸透数据集上进行数值实验, 结果表明该异常检测方法的异常检测性能大幅优于现有方法, 且具有较好的鲁棒性。该研究提出的无监督医学图像异常检测方法和异常区域定位方法可为临床诊疗提供丰富的信息支持, 有望减轻医生的工作量。

**关键词:** 异常检测; 扩散常微分方程; 神经网络; 医学图像; 异常定位

**中图分类号:** TP391

**文献标志码:** A

**文章编号:** 1673-3851(2024)11-0851-10

**引用格式:** 胡显耀, 靳聪明. 基于扩散常微分方程的医学图像异常检测[J]. 浙江理工大学学报(自然科学), 2024, 51(6):851-860.

**Reference Format:** HU Xianyao, JIN Congming. Medical image anomaly detection based on diffusion ordinary differential equations [J]. Journal of Zhejiang Sci-Tech University, 2024, 51(6):851-860.

## Medical image anomaly detection based on diffusion ordinary differential equations

HU Xianyao, JIN Congming

(School of Science, Zhejiang Sci-Tech University, Hangzhou 310018, China)

**Abstract:** To identify abnormal medical images that deviate significantly from normal physiological states, a medical image anomaly detection method, which is based on diffusion ordinary differential equations (ODE) assisted by neural network, is proposed according to the assumption that the relevant features of abnormal images often appear in low-density regions of the feature distribution. Firstly, diffusion ODE is utilized to estimate the likelihood value of the image features; then, a neural network is constructed to fit the likelihood values of the image features at different times estimated by the diffusion ODE; finally, the anomaly score of this method is the weighted average of the likelihood values estimated by the diffusion ODE and the likelihood values estimated by the neural network. Images with high anomaly scores are identified as abnormal images. In addition, an anomaly localization method based on image reconstruction is proposed to determine the abnormal regions of the abnormal images, and the reconstruction errors are calculated to locate the abnormal regions. The numerical experimental results on the BraTS2021 brain tumor dataset and the chest X-ray dataset show that the anomaly detection performance of this method is significantly better than that of other existing methods and has preferable robustness. The approach for unsupervised medical image anomaly detection and the method to locate the

收稿日期: 2024-05-17 网络出版日期: 2024-09-30

基金项目: 国家自然科学基金项目(11571314)

作者简介: 胡显耀(1999—), 男, 安徽合肥人, 硕士研究生, 主要从事深度学习等方面的研究。

通信作者: 靳聪明, E-mail: jincm@zstu.edu.cn

abnormal regions proposed in this article can provide mass information support for clinical diagnosis and treatment, and are expected to reduce the workload of doctors.

**Key words:** anomaly detection; diffusion ordinary differential equations; neural network; medical image; anomaly localization

## 0 引 言

医学图像异常检测是一种在医学图像数据中发现与正常生理状态或指标显著偏离的异常图像的技术,在医疗诊断、疾病预测、患者监护以及药物研发等方面发挥着重要作用。医学图像异常检测的核心在于对海量、高维度的医学图像(如 X 射线图像、MRI 图像等)进行综合分析和模式识别,自动地识别出与正常生理状态或指标不符的异常图像。

通常,标记医学图像需要大量专业知识和时间,因此医学图像异常检测通常考虑无监督图像异常检测方法。无监督图像异常检测方法(下称“异常检测”)主要包括 5 种类型,即基于重构的异常检测<sup>[1-4]</sup>、基于单类分类的异常检测<sup>[5-9]</sup>、基于知识蒸馏的异常检测<sup>[10-14]</sup>、基于记忆库的异常检测<sup>[15-17]</sup>和基于似然的异常检测<sup>[18-20]</sup>。基于重构的异常检测方法<sup>[1-4]</sup>通常训练一个生成模型,例如生成对抗网络<sup>[21]</sup>或扩散模型<sup>[22-24]</sup>等,学习正常数据的分布,并通过训练后的生成模型重构数据,然后通过重构误差来检测异常图像。当使用扩散模型进行异常检测时,在前向扩散过程阶段对图像注入噪声,得到噪声图像;在逆向扩散过程阶段对噪声图像去噪,得到重构图像。如存在较大的重构误差,则表明是异常图像。然而,如果扩散时间较短,重构后的图像与原图像差异很小,不足以识别异常图像;如果扩散时间较长又会破坏图像的结构。因此在构建扩散模型时须尽可能地保留图像结构,使重构误差能反映真实的异常区域。例如,Wyatt 等<sup>[2]</sup>利用具有多尺度单纯形噪声的扩散模型重构图像;Mousakhan 等<sup>[3]</sup>构建了条件扩散模型以获得保留图像结构的重构图像。这些模型利用了扩散模型出色的生成能力,但忽略了扩散模型优质的密度估计性能。基于单类分类的异常检测方法<sup>[5-9]</sup>将正常数据作为一个类别,将不属于这个类别的数据识别为异常。基于知识蒸馏的异常检测方法<sup>[10-14]</sup>利用教师模型和学生模型进行知识传递,在异常图像中通过教师模型和学生模型得到的特征值存在差异,因此该差异值可以检测出异常图像。基于记忆库的异常检测方法<sup>[15-17]</sup>的思想是建立一个包含了正常数据特征的记忆库,检测时计算

样本数据特征与正常数据特征的相似度,相似度较小的图像被识别为异常图像。基于似然的异常检测方法<sup>[18-20]</sup>通常假设异常数据及其特征分布于数据的低密度区域,并根据数据或数据特征的密度值来识别异常图像。

这 5 类异常检测方法都展现了一定的异常检测性能和异常定位性能。在基于重构的异常检测方法中,基于扩散模型的异常检测方法的异常检测性能较差,但可以有效地定位异常区域;基于密度估计的异常检测方法通常采用标准化流模型作为密度估计器,可以得到最好的异常检测性能,但不能较好地定位异常区域。Song 等<sup>[24]</sup>发现扩散模型是一种比标准化流模型性能更优的密度估计器,因此本文考虑用扩散模型作为密度估计器估计图像特征的密度来识别异常图像,并利用扩散模型设计了一种基于图像重构的异常定位方法。

基于似然的异常检测具有较好的异常检测性能,它将图像或图像特征密度值较低的图像识别为异常图像,然而对于模型未学习到的正常图像或图像特征,往往也会得到较低的密度值,这会降低异常检测性能。在有监督异常检测中,分类模型<sup>[25]</sup>通过学习图像或图像特征的正常和异常模式来识别异常图像,然而对于未出现在训练集中的异常模式,其识别能力可能受到显著限制,进而影响检测的全面性和准确性。本文结合这两种异常检测范式,提出了一种神经网络辅助的基于扩散常微分方程(Ordinary differential equations, ODE)的无监督异常检测方法(Anomaly detection with diffusion ODE, AnoDODE)。该异常检测方法分为两个步骤。首先,利用扩散 ODE 估计图像特征的似然值;然后,构建神经网络,本文称之为似然网络,拟合图像特征在不同时刻由扩散 ODE 估计的似然值。通过扩散 ODE 估计的似然值和似然网络估计的似然值的加权平均计算异常分数,异常分数较大的图像被认定为异常图像。如果似然网络估计的似然值可以优化异常分数,那么会提升该方法的整体异常检测性能,似然网络在训练时包含大量正常图像和少量异常图像,这也更符合现实情况。此外,为了确定异常区域,本文提出了一种由原始图像指导的基于

图像重构的异常定位方法,在扩散 ODE 的图像特征去噪过程中引入原始图像作指导,通过解码器重构图像并保留原始图像的结构信息,依据重构误差精确定位及可视化异常区域。

## 1 扩散模型

扩散模型<sup>[22-24]</sup>,又称基于分数的生成模型,在各种图像生成任务中展现了优越的性能,它不仅能够生成高质量样本,而且能够估计数据密度。扩散模型的主要原理是:首先是前向扩散过程,系统地、缓慢地向数据注入噪声,破坏数据的结构,使加噪后的数据满足给定的先验分布,通常是各向同性的高斯分布;然后是学习逆向扩散过程,去噪并生成数据。一旦学习得到参数化的逆向扩散过程,就可以从先验分布中采样,并通过逆向扩散过程迭代生成数据,或者利用连续变量变换定理<sup>[26]</sup>得到数据密度。

前向扩散过程可以由随机微分方程(Stochastic differential equations, SDE)给出,可用微分方程(1)表示:

$$d\mathbf{z}(t) = f(t)\mathbf{z}(t)dt + g(t)d\mathbf{w} \quad (1)$$

其中:  $t \in [0, 1]$  表示时间,  $\mathbf{w}$  是标准维纳过程,  $f(t)$  是  $\mathbf{z}(t)$  的漂移系数,  $g(t)$  是  $\mathbf{z}(t)$  的扩散系数。 $\mathbf{z}(0) \sim p(\mathbf{z}(0))$  是  $t=0$  的初始数据,即数据样本或其特征;通过求解微分方程(1)可以得到加噪后  $t=1$  时的数据  $\mathbf{z}(1)$ , 且  $\mathbf{z}(1) \sim p(\mathbf{z}(1))$ 。要使  $p(\mathbf{z}(1))$  是一个与  $\mathbf{z}(0)$  无关的先验分布,例如具有给定均值和方差的高斯分布,需要设计特殊的函数  $f(t)$  和  $g(t)$ 。

逆向扩散过程可以由 ODE(2)给出:

$$d\mathbf{z} = \left[ f(t)\mathbf{z}(t) - \frac{1}{2}g(t)^2 \nabla_{\mathbf{z}(t)} \log p(\mathbf{z}(t)) \right] dt, \quad t \in [0, 1] \quad (2)$$

其中:  $\mathbf{z}(t) \sim p(\mathbf{z}(t))$ ,  $\nabla_{\mathbf{z}(t)} \log p(\mathbf{z}(t))$  是分布  $p(\mathbf{z}(t))$  的分数。

扩散模型通过训练一个神经网络(称为分数网络)  $\mathbf{s}_\theta(\mathbf{z}(t), t)$  来估计  $\nabla_{\mathbf{z}(t)} \log p(\mathbf{z}(t))$ , 扩散模型的损失函数定义为:

$$\mathbb{E}_{t \sim U(0,1)} [\lambda(t) \mathbb{E}_{\mathbf{z}(0) \sim p(\mathbf{z}(0))} \mathbb{E}_{\mathbf{z}(t) \sim p_{0t}(\mathbf{z}(t) | \mathbf{z}(0))} [\|\mathbf{s}_\theta(\mathbf{z}(t), t) - \nabla_{\mathbf{z}(t)} \log p_{0t}(\mathbf{z}(t) | \mathbf{z}(0))\|_2^2]] \quad (3)$$

其中:  $U(0, 1)$  是  $[0, 1]$  上的均匀分布;  $p_{0t}(\mathbf{z}(t) | \mathbf{z}(0))$  表示从  $\mathbf{z}(0)$  到  $\mathbf{z}(t)$  的条件概率;  $\lambda(t): [0, 1] \rightarrow \mathbf{R}^+$  是一个正权重函数,通常取 1。

在分数网络  $\mathbf{s}_\theta(\mathbf{z}(t), t)$  训练完成后,可以得到

扩散 ODE:

$$d\mathbf{z} = \left[ f(t)\mathbf{z}(t) - \frac{1}{2}g(t)^2 \mathbf{s}_\theta(\mathbf{z}(t), t) \right] dt, \quad t \in [0, 1] \quad (4)$$

于是,从  $\mathbf{z}(1) \sim p(\mathbf{z}(1))$  的样本开始,通过求解扩散 ODE(4),可以得到生成样本  $\mathbf{z}(0)$ 。扩散 ODE(4)是连续标准化流<sup>[26]</sup>的特殊情况。对任意  $\tau \in [0, 1]$ , 可以通过连续变量变换定理<sup>[26]</sup>计算  $\mathbf{z}(\tau)$  的对数似然:

$$\log p(\mathbf{z}(\tau)) = \log p(\mathbf{z}(1)) + \int_\tau^1 \nabla \cdot \tilde{\mathbf{f}}_\theta(\mathbf{z}, t) dt \quad (5)$$

其中:

$$\tilde{\mathbf{f}}_\theta(\mathbf{z}, t) := f(t)\mathbf{z}(t) - \frac{1}{2}g(t)^2 \mathbf{s}_\theta(\mathbf{z}(t), t) \quad (6)$$

$\mathbf{z}(t)$  可以通过求解扩散 ODE(4)获得,  $\nabla \cdot \tilde{\mathbf{f}}_\theta(\mathbf{z}, t)$  通常使用 Skilling-Hutchinson 迹估计器<sup>[27]</sup>估计:

$$\nabla \cdot \tilde{\mathbf{f}}_\theta(\mathbf{z}, t) = \mathbb{E}_{p(\boldsymbol{\varepsilon})} [\boldsymbol{\varepsilon}^\top \nabla \tilde{\mathbf{f}}_\theta(\mathbf{z}, t) \boldsymbol{\varepsilon}] \quad (7)$$

其中:随机变量  $\boldsymbol{\varepsilon}$  满足  $\mathbb{E}_{p(\boldsymbol{\varepsilon})} [\boldsymbol{\varepsilon}] = 0$ ,  $\text{Cov}_{p(\boldsymbol{\varepsilon})} [\boldsymbol{\varepsilon}] = \mathbf{I}$ ,  $\nabla \tilde{\mathbf{f}}_\theta$  表示  $\tilde{\mathbf{f}}_\theta(\cdot, t)$  的雅可比行列式。向量雅可比积  $\boldsymbol{\varepsilon}^\top \nabla \tilde{\mathbf{f}}_\theta(\mathbf{z}, t)$  可以使用反向模式自动微分进行有效计算。因此,可以对  $\boldsymbol{\varepsilon} \sim p(\boldsymbol{\varepsilon})$  进行采样,然后使用  $\boldsymbol{\varepsilon}^\top \nabla \tilde{\mathbf{f}}_\theta(\mathbf{z}, t) \boldsymbol{\varepsilon}$  计算  $\nabla \cdot \tilde{\mathbf{f}}_\theta(\mathbf{z}, t)$  的有效无偏估计。

密度估计器的性能通常用 BPD (Bits per dim) 分数评价。BPD 分数是每个维度的比特数,可以通过计算数据或数据特征的负对数似然并更改对数基数得到:

$$f_{\text{bpd}}(\mathbf{z}(t), t) = -\log p(\mathbf{z}(t)) / (\log 2 \cdot \prod_{i=1}^3 d_i), \quad t \in [0, 1] \quad (8)$$

其中:  $d_1, d_2, d_3$  分别表示数据或数据特征的高度、宽度和通道数。

分类器指导的扩散模型<sup>[28]</sup>通过额外训练一个分类器,利用条件信息  $y$  指导预训练好的扩散模型生成与给定条件信息  $y$  相关联的数据样本。分类器指导的扩散模型在逆向扩散过程(2)中的分数为:

$$\nabla_{\mathbf{z}(t)} \log p(\mathbf{z}(t) | y) = \nabla_{\mathbf{z}(t)} \log p(\mathbf{z}(t)) + \gamma \nabla_{\mathbf{z}(t)} \log p(y | \mathbf{z}(t)) \quad (9)$$

其中:分类器  $p(y | \mathbf{z}(t))$  需要额外训练,  $\gamma$  为超参数。当  $\gamma=0$  时,扩散模型会完全忽略条件信息;而随着  $\gamma$  增大,扩散模型会生成越来越符合条件信息的样本。

## 2 本文模型

本文提出的 AnoDODE 方法由两个步骤组成。首先,利用扩散 ODE 估计图像特征的似然值;然后,构建似然网络拟合图像特征在扩散 ODE 模型中不同时刻的似然值。AnoDODE 方法的异常分数是扩散 ODE 估计的似然值和似然网络估计的似然值的加权平均。此外,由于扩散 ODE 模型可以生成图像特征,本文提供了一种基于重构的异常定位方法,提供异常区域的可视化,可以辅助用户对异常

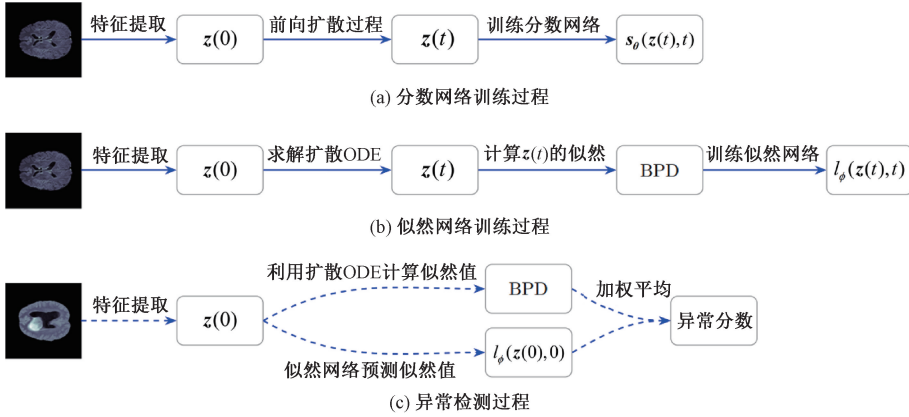


图1 AnoDODE 方法的流程图

前向扩散过程的噪声注入方法采用 VPSDE (Variance preserving stochastic differential equation)<sup>[24]</sup>。VPSDE 的前向扩散过程为:

$$d\mathbf{z}(t) = -\frac{1}{2}\beta(t)\mathbf{z}(t)dt + \sqrt{\beta(t)}d\mathbf{w} \quad (10)$$

其中:  $\beta(t) = \bar{\beta}_{\min} + t(\bar{\beta}_{\max} - \bar{\beta}_{\min})$ ,  $t \in [0, 1]$ ,  $\bar{\beta}_{\min} = 0.1$ ,  $\bar{\beta}_{\max} = 20$ 。由于 VPSDE 具有仿射漂移系数,因此转移概率  $p_{\text{Or}}(\mathbf{z}(t) | \mathbf{z}(0))$  为高斯分布且封闭形式为:

$$\begin{aligned} p_{\text{Or}}(\mathbf{z}(t) | \mathbf{z}(0)) &= \mathcal{N}\left(\mathbf{z}(t); \mathbf{z}(0)e^{-\frac{1}{2}\int_0^t \beta(s)ds}, \mathbf{I} - \mathbf{I}e^{-\int_0^t \beta(s)ds}\right) \\ &= \mathcal{N}\left(\mathbf{z}(t); \mathbf{z}(0)e^{-\frac{1}{4}t^2(\bar{\beta}_{\max} - \bar{\beta}_{\min}) - \frac{1}{2}t\bar{\beta}_{\min}}, \right. \\ &\quad \left. \mathbf{I} - \mathbf{I}e^{-\frac{1}{2}t^2(\bar{\beta}_{\max} - \bar{\beta}_{\min}) - t\bar{\beta}_{\min}}\right), \quad t \in [0, 1] \end{aligned} \quad (11)$$

于是分数网络  $s_{\theta}(\mathbf{z}(t), t)$  可以通过极小化损失函数式(3)进行训练。在异常检测中,利用方程(10)对应的逆向扩散 ODE,即:

$$d\mathbf{z}(t) = \left\{ -\frac{1}{2}\beta(t)\mathbf{z}(t) - \frac{1}{2}\beta(t)s_{\theta}(\mathbf{z}(t), t) \right\} dt, \quad t \in [0, 1] \quad (12)$$

得到中间时刻的图像特征  $\mathbf{z}(t)$ , 并通过连续变量变换公式(5)得到图像特征  $\mathbf{z}(0)$  的对数似然值,然后

区域进行定位。

### 2.1 扩散过程

扩散模型的训练集仅为正常数据。在训练过程中,利用预训练的特征编码器,即通过使用大规模数据集进行预训练的神经网络模型,提取医学图像的特征  $\mathbf{z}(0)$ 。然后,利用前向扩散过程得到注入噪声的图像特征  $\mathbf{z}(t)$ , 并训练分数网络  $s_{\theta}(\mathbf{z}(t), t)$  拟合图像特征分布的分数  $\nabla_{\mathbf{z}(t)} \log p(\mathbf{z}(t))$ , 以便在异常检测中利用扩散 ODE 对图像特征进行密度估计。图 1(a)展示了分数网络训练过程。

通过式(8)得到的图像特征  $\mathbf{z}(0)$  的 BPD 分数作为异常分数。

### 2.2 似然网络

似然网络的训练集包含正常和异常数据。在训练中,利用训练好的分数网络  $s_{\theta}(\mathbf{z}(t), t)$  指导似然网络  $l_{\phi}(\mathbf{z}(t), t)$  的训练。似然网络的训练步骤为:第一步,利用预训练的特征编码器得到医学图像特征  $\mathbf{z}(0)$ 。第二步,通过求解扩散 ODE(12)可以得到任意中间时刻  $t \in [0, 1]$  的图像特征  $\mathbf{z}(t)$ 。第三步,利用连续变量变换公式(5)获得图像特征  $\mathbf{z}(t)$  的似然值,然后通过式(8)得到图像特征  $\mathbf{z}(t)$  的 BPD 分数。第四步,构造一个嵌入时间  $t$  的似然网络  $l_{\phi}(\mathbf{z}(t), t)$  拟合这些中间时刻图像特征的 BPD 分数,损失函数定义为:

$$\mathbb{E}_{t \sim U[0, 1]} [\|f_{\text{bpd}}(\mathbf{z}(t), t) - l_{\phi}(\mathbf{z}(t), t)\|_2^2] \quad (13)$$

其中:  $U[0, 1]$  是均匀分布,  $f_{\text{bpd}}(\mathbf{z}(t), t)$  是第三步估计的图像特征  $\mathbf{z}(t)$  的 BPD 分数。图 1(b)展示了似然网络的训练过程。在异常检测时,异常分数是似然网络生成的 BPD 分数,即  $l_{\phi}(\mathbf{z}(0), 0)$ 。直接训练似然网络拟合  $\mathbf{z}(0)$  的 BPD 分数会出现过拟合现象,而构造的嵌入时间  $t$  的似然网络训练方法可



以有效缓解过拟合现象。

### 2.3 异常分数

由于前向扩散过程是在正常图像特征上训练的,因此在异常检测时,正常图像特征会得到较低的 BPD 分数;而异常图像特征会得到较高的 BPD 分数,即异常图像特征出现在分布的低密度区域。似然网络在正常和异常图像特征上训练,可以学习到图像特征的正常和异常模式,对于异常图像特征,会得到较高的 BPD 分数。本文提出的 AnoDODE 方法结合了这两部分得到的 BPD 分数,即异常分数是扩散 ODE 估计的似然值和似然网络估计的似然值的加权平均:

$$S = \omega \cdot l_{\phi}(\mathbf{z}(0), 0) + (1 - \omega) \cdot f_{\text{bpd}}(\mathbf{z}(0), 0) \quad (14)$$

其中:  $\omega$  为权重,  $l_{\phi}(\mathbf{z}(0), 0)$  是似然网络生成的 BPD 分数,  $f_{\text{bpd}}(\mathbf{z}(0), 0)$  是利用扩散 ODE 得到的 BPD 分数。图 1(c) 展示了异常检测过程。利用扩散 ODE 估计似然值需要迭代求解 ODE(12), 这需要大量的异常检测时间,而似然网络训练好后可以一步生成图像的似然值,因此对于在线即时应用,可以直接将模型的异常分数设置为似然网络的似然值,即  $\omega = 1$ 。

### 2.4 异常定位

在异常定位时,对医学图像  $x$  中提取的图像特征  $\mathbf{z}(0)$  加噪到某个中间时刻  $\tau \in [0, 1]$  得到加噪后的图像特征  $\mathbf{z}(\tau)$ , 即  $\mathbf{z}(0) \rightarrow \mathbf{z}(\tau)$ ; 然后将其去噪即可获得重构的图像特征, 记为  $\mathbf{z}'(0)$ , 即  $\mathbf{z}(\tau) \rightarrow \mathbf{z}'(0)$ 。此外,使用正常图像,训练一个解码器将重构特征  $\mathbf{z}'(0)$  重构成图像,即  $\mathbf{z}'(0) \rightarrow \mathbf{x}'$ 。由于前向扩散过程和解码器是用正常图像训练的,因此异常图像在重构后会消除异常信息,使得重构图像与原始异常图像显著不同。重构图像  $\mathbf{x}'$  和原始图像  $\mathbf{x}$  之间的均方误差  $\|\mathbf{x}' - \mathbf{x}\|_2^2$  为图像中的异常定位提供了有价值的信息,重构误差较大的区域即为异常区域。需要注意的是,中间时刻  $\tau$  的选取会影响异常定位性能。如果  $\tau$  较小(接近 0),加噪过程可能无法破坏图像特征中的异常信息,因而难以生成正常图像的特征;相反,如果  $\tau$  较大(接近 1),加噪过程可能会损坏图像特征中的正常部分的信息,使重构图像无法保留原始图像的结构。因此,本文考虑引入原始图像  $\mathbf{x}$  作为标签指导去噪过程,生成符合原始图像结构的图像特征。类似式(9),本文提出的异常定位方法在去噪过程的新分数为:

$$\nabla \log p(\mathbf{z}(t) | \mathbf{x}) = \nabla \log p(\mathbf{z}(t)) - \gamma \nabla \log \|\mathbf{x} - D_{\eta}(\mathbf{z}(t), t)\|_2^2 \quad (15)$$

其中:  $\gamma$  为超参数,  $D_{\eta}(\mathbf{z}(t), t)$  为标签  $\mathbf{x}$  的预测器,需要额外训练。在训练预测器时,  $\mathbf{z}(t)$  由求解前向扩散过程(10)得到。显然,当  $\gamma = 0$  时,扩散模型会完全忽略标签信息  $\mathbf{x}$ ;而  $\gamma$  越大,扩散模型会生成与原始图像结构越相似的样本。利用新的去噪方法重构图像特征后,将  $D_{\eta}(\cdot, 0)$  作为解码器就可得到重构图像。

## 3 数值实验

### 3.1 医学图像数据集

BraTS2021<sup>[29]</sup>是 2021 年脑肿瘤分割挑战赛的数据集。该数据集由 3D 多模态磁共振图像(MRI)组成。在实验中使用的是从这些 3D MRI 图像中提取的深度范围在 60 到 100 之间的 2D 轴向切片<sup>[30]</sup>。训练集包含 7500 个健康切片,验证集包含 44 个肿瘤或异常切片和 39 个健康切片,测试集包含 3075 个肿瘤切片和 640 个健康切片。

X 射线胸透(Chest X-Ray)图像数据集<sup>[25]</sup>是一个公开的医学影像数据集,收集并标记了 5856 名儿童的 X 射线胸透图像,用于训练异常检测模型来识别肺炎图像。在实验中,训练集包含 1349 张正常图像,验证集包含 32 张诊断为肺炎的异常图像(17 张细菌肺炎图像和 15 张病毒肺炎图像)和 16 张正常图像,测试集包含 390 张诊断为肺炎的异常图像(242 张细菌肺炎图像和 148 张病毒肺炎图像)和 234 张正常图像。

所有对比实验都采用了相同的训练数据和测试数据。由于似然网络需要学习异常图像的异常模式,在实验中,似然网络的训练数据是验证集中的无标签数据。

### 3.2 实验设置

分数网络、似然网络和解码器与文献[31]采用相同的架构,都通过正弦嵌入的方式嵌入时间  $t$ 。所有实验的特征提取器采用 EfficientNet-B5<sup>[32]</sup>第 36 层的输出,它将图像转化为具有 304 通道的特征。EfficientNet-B5 的参数在实验期间是固定的,采用在 ImageNet 数据集<sup>[33]</sup>上进行预训练的结果。模型采用 RK45 ODE 求解器用于计算不同时刻的图像特征及其对数似然。在异常定位时,图像特征的去噪过程中采样步数设置为 500。模型训练采用了 Adam 优化器<sup>[34]</sup>,在所有实验中,学习率均设置为  $10^{-5}$ 。异常分数中的权重  $\omega$  设置为 0.6。

3.3 医学图像的异常检测

本文将提出的 AnoDODE 方法应用于两个医学图像数据集上,实验结果与 7 种基于深度学习的方法的结果进行了比较,其中:2 种为基于重构的方法 (Autoencoder 和 f-AnoGAN),1 种为基于单类分类的方法 (Cutpaste),3 种为基于知识蒸馏的方法 (MKD、RD4AD 和 EfficientAD),1 种为基于似然的方法 (CS-Flow)。本文采用多种指标对

这些方法进行评价,包括 AUC 分数、F1 分数和 ACC 值,其中用于计算 ACC 值的阈值根据最佳 F1 分数确定。表 1 展示了在 BraTS2021 数据集和 X 射线胸透数据集上的实验结果。图 2 展示了所有模型在 BraTS2021 数据集和 X 射线胸透数据集上对应的 ROC 曲线,提出的 AnoDODE 方法在两个医学数据集上都获得了最佳 AUC 分数。

表 1 不同方法在 BraTS2021 数据集和 X 射线胸透数据集上的实验评价指标%

方法	BraTS2021 数据集评价指标			X 射线胸透数据集评价指标		
	AUC 分数	F1 分数	ACC 值	AUC 分数	F1 分数	ACC 值
Autoencoder	65.54	90.64	82.93	79.32	80.57	71.63
f-AnoGAN <sup>[1]</sup>	80.96	91.63	85.06	65.36	76.80	62.34
Cutpaste <sup>[11]</sup>	79.17	91.06	83.93	62.92	76.80	62.34
MKD <sup>[13]</sup>	81.26	92.02	85.84	90.25	86.15	83.81
RD4AD <sup>[14]</sup>	89.34	92.64	87.24	79.01	82.69	76.44
EfficientAD <sup>[17]</sup>	80.45	91.11	84.19	76.09	84.12	79.01
CS-Flow <sup>[18]</sup>	90.88	92.60	87.51	91.24	86.59	82.53
AnoDODE(本文)	93.26	93.54	89.39	94.52	91.07	88.94

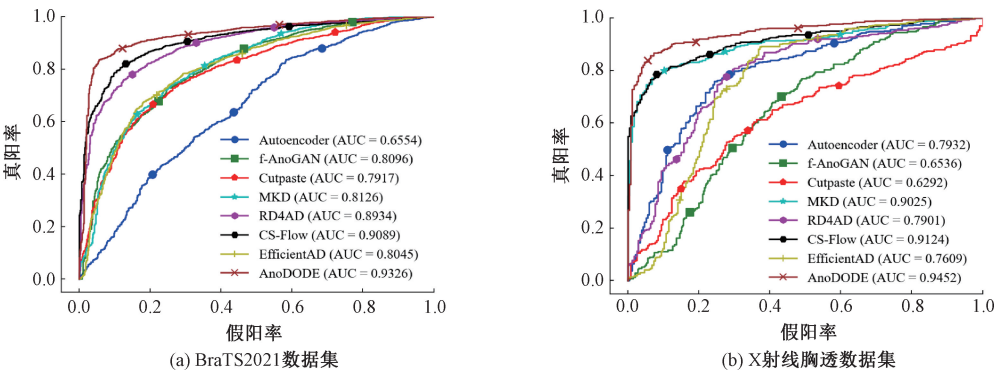


图 2 所有方法在 BraTS2021 数据集和 X 射线胸透数据集上对应的 ROC 曲线

在 BraTS2021 数据集上,AnoDODE 方法在所有指标上取得了最优性能。值得注意的是,Autoencoder 方法表现出的性能最差,但由于测试数据集中正负样本分布不平衡,它得到了看似不错的 F1 分数和 ACC 值。f-AnoGAN 的性能符合预期,结果优于 Autoencoder 方法。Cutpaste 没有取得良好的结果,可能是因为它合成异常图像的方式不适合 BraTS2021 数据集。MKD 和 RD4AD 是基于知识蒸馏的异常检测方法,在 RD4AD 在 BraTS2021 数据集上表现良好,而 MKD 和 EfficientAD 表现较差。CS-Flow 基于图像特征的密度值进行异常检测,取得了较好的异常检测性能。然而,本文提出的神经网络辅助的基于扩散 ODE 的异常检测方法在所有指标上大幅优于 CS-Flow。

在 X 射线胸透数据集上,AnoDODE 方法在所有指标上都大幅优于其他代表性方法。f-AnoGAN

和 Cutpaste 方法的异常检测性能较差,并且低于 Autoencoder 方法,这可能是因为训练数据不足导致的。同时,实验发现 RD4AD 和 EfficientAD 方法的异常检测性能比 MKD 方法差,这说明基于知识蒸馏的异常检测方法存在不稳定性,也表明这类方法存在缺乏可解释性的缺点。

为了进一步验证 AnoDODE 方法的有效性,图 3 给出了异常分数的直方图,其中正常和异常数据用不同灰度区分,横轴表示异常分数,纵轴表示频率和组距的比值。结果表明,异常数据的异常分数通常大于正常数据的异常分数,而重叠面积仅占总面积的一小部分,这表明所提出的异常检测方法能有效地区分正常和异常样本。此外,图 4 中异常检测分类结果的混淆矩阵表明,所提出的模型可以对正常和异常图像进行有效分类。

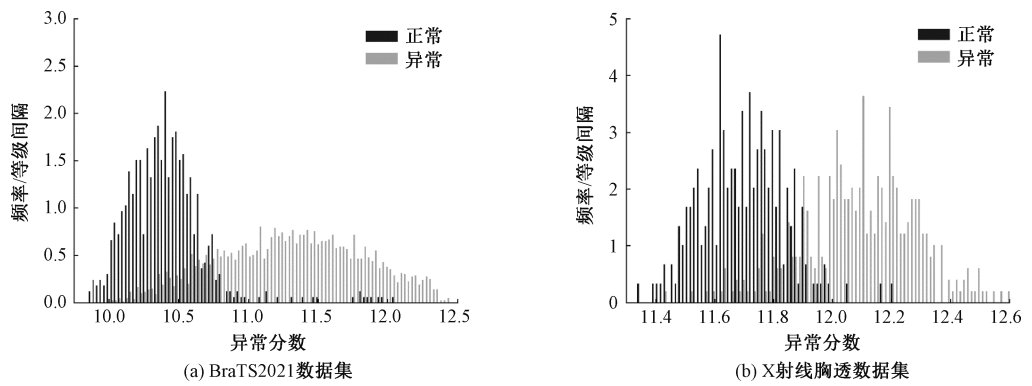


图 3 AnoDODE 方法在 BraTS2021 数据集和 X 射线胸透数据集上的异常分数分布直方图

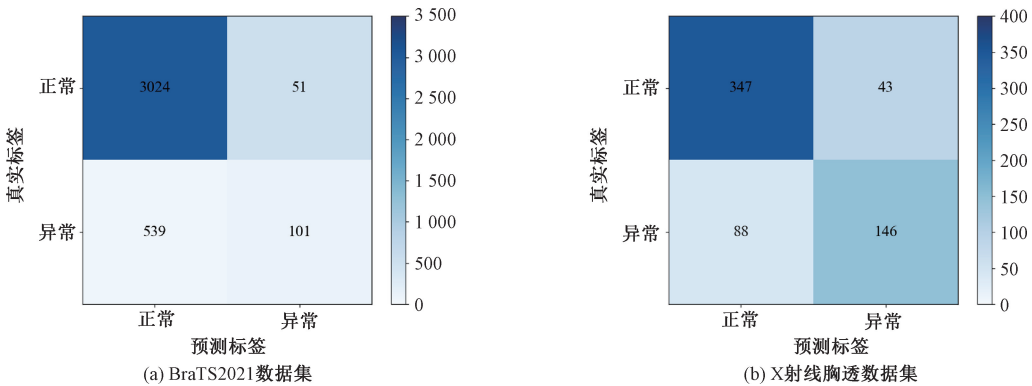


图 4 AnoDODE 方法在 BraTS2021 数据集和 X 射线胸透数据集上的混淆矩阵

3.4 医学图像的异常定位

本文采用原始图像指导的基于重构的异常定位方法进行医学图像异常病变区域的定位,其中扩散时间 $\tau$ 设定为 0.5,缩放因子 $\gamma$ 设定为 20。图 5 展示了在 BraTS2021 数据集和 X 射线胸透数据集上异常定位的实验结果,每张图像包含原始图像(第 1 列)、重构图像(第 2 列)和热图 $\|\mathbf{x}' - \mathbf{x}\|_2^2$ (第 3 列)。

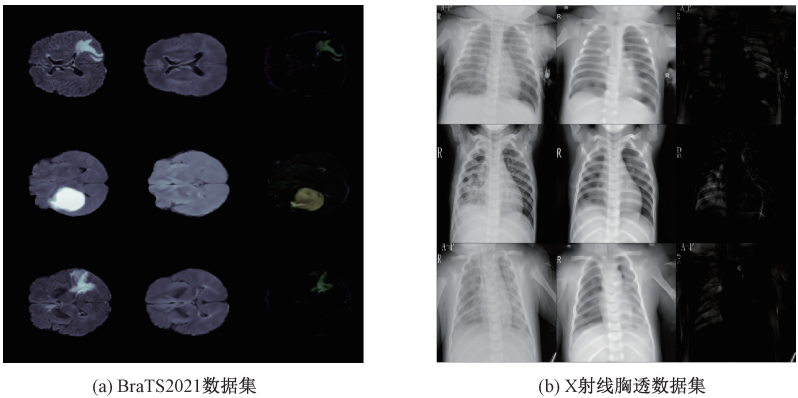


图 5 BraTS2021 数据集和 X 射线胸透数据集中异常图像的异常定位

3.5 消融实验

在 AnoDODE 方法中,式(14)中的异常分数包含权参数 $\omega$ ,本文分析了不同 $\omega$ 对异常检测性能的影响。图 6 展示了 AnoDODE 方法在 BraTS2021 数据集和 X 射线胸透数据集上的实验结果随 $\omega$ 的变化曲线。当 $\omega = 0$ 时,模型的异常分

在 BraTS2021 数据集上,提出的异常定位方法较好地保留了脑切片的轮廓,重构出消除脑肿瘤的脑切片重构图像,重构误差有助于精准地识别脑肿瘤。在 X 射线胸透数据集上,提出的异常定位方法同样可以保留原始图像的胸部轮廓,从而使重构误差可以精准地识别肺炎区域。本文未进一步优化提出的异常定位方法,仅对异常区域进行了可视化。

数就是扩散 ODE 得到的图像特征的 BPD 分数;而当 $\omega = 1$ 时,模型的异常分数就是似然网络估计的图像特征的 BPD 分数;实验表明,当 $\omega = 0.6$ 左右时,模型取得最优结果,这表明了似然网络的引入优化了模型的异常分数,有助于提升模型的异常检测性能。



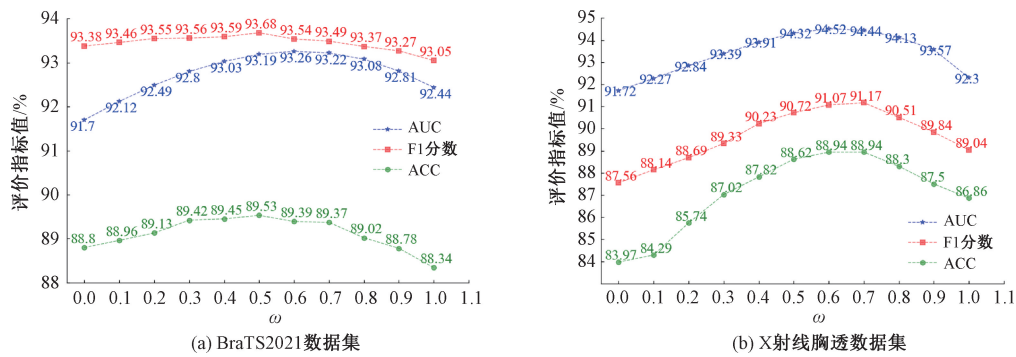


图 6 AnoDODE 方法在 BraTS2021 数据集和 X 射线胸透数据集上的实验结果随  $\omega$  的变化曲线

3.6 可靠性研究

为了研究 AnoDODE 方法的鲁棒性,固定训练好的似然网络和分数网络,在异常检测过程中,向图像  $x$  添加少量的噪声:  $\bar{x} = x + 0.1\epsilon$ , 其中  $\epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ ,  $\bar{x}$  为加噪后的图像。表 2 展示了 BraTS2021 数据集和 X 射线胸透数据集加噪的测试图像和原测试图像的实验结果对比。实验结果表明,在图像加噪后,模型会损失一部分异常检测性能。当  $\omega = 0$  时,模型的异常分数仅由扩散

ODE 模型求得,损失较多异常检测性能,对于 X 射线胸透数据集尤为明显;当  $\omega = 1$  时,模型的异常分数由似然网络求得,损失少量异常检测性能,依旧保持良好的性能;当  $\omega = 0.6$  时,模型有利于保持良好的异常检测性能,在 BraTS2021 数据集上实现最优性能,在 X 射线胸透数据集上实现较优的性能。实验结果表明,提出的似然网络辅助的基于扩散 ODE 的异常检测方法具有较好的鲁棒性。

表 2 BraTS2021 数据集和 X 射线胸透数据集原测试图像  $x$  和加噪的测试图像  $\bar{x}$  的实验评价指标 %

方法	BraTS2021 数据集评价指标			X 射线胸透数据集评价指标		
	AUC 分数	F1 分数	ACC 值	AUC 分数	F1 分数	ACC 值
$x - \text{AnoDODE}(\omega = 0)$	91.70	93.38	88.80	91.72	87.56	83.93
$x - \text{AnoDODE}(\omega = 1)$	92.44	93.05	88.34	92.30	89.04	86.86
$x - \text{AnoDODE}(\omega = 0.6)$	93.26	93.54	89.39	94.52	91.07	88.94
$\bar{x} - \text{AnoDODE}(\omega = 0)$	90.38	92.98	87.73	82.11	82.30	76.28
$\bar{x} - \text{AnoDODE}(\omega = 1)$	92.20	92.98	88.21	92.11	88.75	86.06
$\bar{x} - \text{AnoDODE}(\omega = 0.6)$	92.90	93.26	88.75	92.28	88.18	85.74

4 结 论

本文提出了一种神经网络辅助的基于扩散 ODE 的无监督医学图像异常检测方法,以及一种由原始图像指导的基于图像重构的异常定位方法。在两个医学图像数据集上的实验结果表明,本文提出的 AnoDODE 方法是有效的和鲁棒的,与基于重构的异常检测方法、基于知识蒸馏的异常检测方法等已有方法相比,异常检测性能有很大提升;本文提出的异常定位方法可以有效地定位医学图像病变区域。该方法可以在医学图像标签缺少的情况下,提供一种基于密度估计的无监督异常检测方案,较好地识别异常医学图像,并定位出异常病变区域,为临床诊疗提供更丰富的信息支持,并减轻医生的工作量。同时,该方法也可应用于其他医学图像和工业图像异常检测任务。后续研究可考虑利用扩散 ODE 的最新技术进一步提升模型的异常检测性能。

参考文献:

[1] Schlegl T, Seeböck P, Waldstein S M, et al. F-AnoGAN: Fast unsupervised anomaly detection with generative adversarial networks [J]. Medical Image Analysis, 2019, 54: 30-44.

[2] Wyatt J, Leach A, Schmon S M, et al. AnoDDPM: Anomaly detection with denoising diffusion probabilistic models using simplex noise [C] // 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). New Orleans, LA, USA. IEEE, 2022: 649-655.

[3] Mousakhan A, Brox T, Tayyub J. Anomaly detection with conditioned denoising diffusion models[EB/OL]. (2023-12-03) [2024-07-01]. <http://arxiv.org/abs/2305.15956>.

[4] Behrendt F, Bhattacharya D, Krüger J, et al. Patched diffusion models for unsupervised anomaly detection in



- brain MRI[C]//Medical Imaging with Deep Learning. Paris, France: PMLR, 2024: 1019-1032.
- [5] Schölkopf B, Platt J C, Shawe-Taylor J, et al. Estimating the support of a high-dimensional distribution[J]. *Neural Computation*, 2001, 13(7): 1443-1471.
- [6] Tax D M J, Duin R P W. Support vector data description[J]. *Machine Learning*, 2004, 54(1): 45-66.
- [7] Ruff L, Vandermeulen R, Goernitz N, et al. Deep one-class classification [C] // International Conference on Machine Learning. Stockholmsmässan, Stockholm Sweden. PMLR, 2018: 4393-4402.
- [8] Li C L, Sohn K, Yoon J, et al. CutPaste: Self-supervised learning for anomaly detection and localization[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Nashville, TN, USA. IEEE, 2021: 9659-9669.
- [9] Liu Z K, Zhou Y M, Xu Y S, et al. SimpleNet: A simple network for image anomaly detection and localization [C] // 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Vancouver, BC, Canada. IEEE, 2023: 20402-20411.
- [10] 邢鹏, 蒋鑫, 潘永华, 等. 基于特征约束蒸馏学习的视觉异常检测[J]. *软件学报*, 2023, 34(9): 4378-4391.
- [11] Salehi M, Sadjadi N, Baselizadeh S, et al. Multiresolution knowledge distillation for anomaly detection [C] // 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Nashville, TN, USA. IEEE, 2021: 14897-14907.
- [12] Deng H Q, Li X Y. Anomaly detection via reverse distillation from one-class embedding[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans, LA, USA. IEEE, 2022: 9727-9736.
- [13] Batzner K, Heckler L, König R. EfficientAD: Accurate visual anomaly detection at millisecond-level latencies[C]//2024 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). Waikoloa, HI, USA. IEEE, 2024: 127-137.
- [14] Rudolph M, Wehrbein T, Rosenhahn B, et al. Asymmetric student-teacher networks for industrial anomaly detection [C] // 2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). Waikoloa, HI, USA. IEEE, 2023: 2591-2601.
- [15] Defard T, Setkov A, Loesch A, et al. PaDiM: A patch distribution modeling framework for anomaly detection and localization [C] // Pattern Recognition. ICPR International Workshops and Challenges (ICPR 2021). Cham: Springer, 2021: 475-489.
- [16] Roth K, Pemula L, Zepeda J, et al. Towards total recall in industrial anomaly detection[C]// 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans, LA, USA. IEEE, 2022: 14298-14308.
- [17] Hyun J, Kim S, Jeon G, et al. ReConPatch: Contrastive patch representation learning for industrial anomaly detection [C] // 2024 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). Waikoloa, HI, USA. IEEE, 2024: 2041-2050.
- [18] Rudolph M, Wandt B, Rosenhahn B. Same same but DifferNet: Semi-supervised defect detection with normalizing flows[C]//2021 IEEE Winter Conference on Applications of Computer Vision (WACV). Waikoloa, HI, USA. IEEE, 2021: 1906-1915.
- [19] Rudolph M, Wehrbein T, Rosenhahn B, et al. Fully convolutional cross-scale-flows for image-based defect detection[C]//2022 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). Waikoloa, HI, USA. IEEE, 2022: 1829-1838.
- [20] Zhao Y, Ding Q, Zhang X. AE-FLOW: Autoencoders with normalizing flows for medical images anomaly detection [C/OL] // The Eleventh International Conference on Learning Representations. ICLR 2023. (2023-02-28) [2024-06-12]. <https://openreview.net/forum?id=9OmCr1q54Z>.
- [21] Creswell A, White T, Dumoulin V, et al. Generative adversarial networks: An overview[J]. *IEEE Signal Processing Magazine*, 2018, 35(1): 53-65.
- [22] Sohl-Dickstein J, Weiss E, Maheswaranathan N, et al. Deep unsupervised learning using nonequilibrium thermodynamics[C]//International Conference on Machine Learning. Lille, France: PMLR, 2015: 2256-2265.
- [23] Ho J, Jain A, Abbeel P. Denoising diffusion probabilistic models[J]. *Advances in Neural Information Processing Systems*, 2020, 33: 6840-6851.
- [24] Song Y, Sohl-Dickstein J, Kingma D P, et al. Score-based generative modeling through stochastic differential equations [EB/OL]. (2021-02-10) [2024-07-01]. <http://arxiv.org/abs/2011.13456>.
- [25] Kermany D S, Goldbaum M, Cai W, et al. Identifying medical diagnoses and treatable diseases by image-based deep learning[J]. *Cell*, 2018, 172(5): 1122-1131.
- [26] Chen R T Q, Rubanova Y, Bettencourt J, et al. Neural ordinary differential equations[J]. *Advances in Neural Information Processing Systems*, 2018, 31:

6571-6583.

[27] Grathwohl W, Chen R T Q, Bettencourt J, et al. FFJORD: Free-form continuous dynamics for scalable reversible generative models[EB/OL]. (2018-10-22) [2024-07-01]. <http://arxiv.org/abs/1810.01367>.

[28] Dhariwal P, Nichol A. Diffusion models beat gans on image synthesis[J]. Advances in Neural Information Processing Systems, 2021, 34: 8780-8794.

[29] Baid U, Ghodasara S, Mohan S, et al. The RSNA-ASNR-MICCAI BraTS 2021 benchmark on brain tumor segmentation and radiogenomic classification[EB/OL]. (2021-09-12) [2024-07-01]. <http://arxiv.org/abs/2107.02314>.

[30] Bao J, Sun H, Deng H, et al. Bmad: Benchmarks for medical anomaly detection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, WA, USA. IEEE, 2024: 4042-4053.

[31] Rombach R, Blattmann A, Lorenz D, et al. High-resolution image synthesis with latent diffusion models [C]// 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans, LA, USA. IEEE, 2022: 10674-10685.

[32] Tan M X, Le Q V. EfficientNet: Rethinking model scaling for convolutional neural networks [C] // International Conference on Machine Learning. Long Beach, CA, USA. PMLR, 2019: 6105-6114.

[33] Deng J, Dong W, Socher R, et al. ImageNet: A large-scale hierarchical image database [C] // 2009 IEEE Conference on Computer Vision and Pattern Recognition. Miami, FL, USA. IEEE, 2009: 248-255.

[34] Kingma D P, Ba J. Adam: A method for stochastic optimization[EB/OL]. (2017-01-30) [2024-07-01]. <http://arxiv.org/abs/1412.6980>.

(责任编辑:康 锋)