



## 基于ODE扩散模型的多类异常检测和定位

蒋世杰, 夏秀山, 翟伟, 曹洋

引用本文:

蒋世杰, 夏秀山, 翟伟, 等. 基于ODE扩散模型的多类异常检测和定位[J]. 智能系统学报, 2025, 20(2): 376-388.

JIANG Shijie, XIA Xiushan, ZHAI Wei, et al. ODE diffusion model for multiclass anomaly detection and localization[J]. *CAAI Transactions on Intelligent Systems*, 2025, 20(2): 376-388.

在线阅读 View online: <https://dx.doi.org/10.11992/tis.202402022>

## 您可能感兴趣的其他文章

### 多条件多样本RNA-Seq数据的剪切异构体表达水平估计

Estimation of transcription variant expression level based on multi-condition multi-sample RNA-Seq data

智能系统学报. 2021, 16(6): 1126-1135 <https://dx.doi.org/10.11992/tis.202101028>

### 基于互信息的多块k近邻故障监测及诊断

Multiblock  $k$ -nearest neighbor fault monitoring and diagnosis based on mutual information

智能系统学报. 2021, 16(4): 717-728 <https://dx.doi.org/10.11992/tis.202007035>

### 利用置信规则库构建WSN节点故障检测模型

Constructing a WSN node fault detection model using the belief rule base

智能系统学报. 2021, 16(3): 511-517 <https://dx.doi.org/10.11992/tis.202009006>

### 移动通信网络的中性集故障诊断方法研究

Research on neutral set fault diagnosis method for mobile communication networks

智能系统学报. 2020, 15(5): 864-869 <https://dx.doi.org/10.11992/tis.201906031>

### 一种双优选的半监督回归算法

A dual-optimal semi-supervised regression algorithm

智能系统学报. 2019, 14(4): 689-696 <https://dx.doi.org/10.11992/tis.201805010>

### 中医脉诊信号的无监督聚类分析研究

Unsupervised clustering analysis of human-pulse signal in traditional Chinese medicine

智能系统学报. 2018, 13(4): 564-570 <https://dx.doi.org/10.11992/tis.201703030>

DOI: 10.11992/tis.202402022

网络出版地址: <https://link.cnki.net/urlid/23.1538.tp.20250108.0939.019>

# 基于 ODE 扩散模型的多类异常检测和定位

蒋世杰<sup>1</sup>, 夏秀山<sup>2</sup>, 翟伟<sup>1</sup>, 曹洋<sup>1</sup>

(1. 中国科学技术大学 自动化系, 安徽 合肥 230027; 2. 中国科学技术大学 先进技术研究院, 安徽 合肥 230031)

**摘要:** 多类异常检测和定位旨在训练一个单一模型, 在多种场景下该模型能够识别出偏离正常的异常区域。最近基于扩散模型的方法在该任务中表现出色而开始受到关注, 然而, 现有方法侧重于改进扩散模型去噪网络, 通过添加更多约束, 保持多步生成的高一致性, 实现更高的重构性能, 但更多的采样步数也意味着更高的计算开销。为此, 本文提出了一种基于常微分方程 (ordinary differential equations, ODE) 扩散模型的多类异常检测和定位方法, 只需一步即可实现高质量的重构生成, 同时引入时间步感知网络来缓解采样步数少可能导致的一致性和恒等捷径问题, 从而进一步提高重构质量。在通用的基准数据集 MVTec-AD 上进行的实验结果表明, 本文方法在精度上可与当前最先进方法相媲美, 但是计算量更低速度更快, 满足了工业异常检测和定位的高精度和实时性需求。

**关键词:** 缺陷检测; 异常检测; 异常定位; 扩散模型; 去噪网络; 常微分方程; 无监督学习; 时间步感知

**中图分类号:** TP391    **文献标志码:** A    **文章编号:** 1673-4785(2025)02-0376-13

中文引用格式: 蒋世杰, 夏秀山, 翟伟, 等. 基于 ODE 扩散模型的多类异常检测和定位 [J]. 智能系统学报, 2025, 20(2): 376-388.

英文引用格式: JIANG Shijie, XIA Xiushan, ZHAI Wei, et al. ODE diffusion model for multiclass anomaly detection and localization[J]. CAAI transactions on intelligent systems, 2025, 20(2): 376-388.

## ODE diffusion model for multiclass anomaly detection and localization

JIANG Shijie<sup>1</sup>, XIA Xiushan<sup>2</sup>, ZHAI Wei<sup>1</sup>, CAO Yang<sup>1</sup>

(1. Department of Automation, University of Science &amp; Technology of China, Hefei 230027, China; 2. Institute of Advanced Technology, University of Science &amp; Technology of China, Hefei 230031, China)

**Abstract:** Multiclass anomaly detection and localization methods aim to train a single model capable of identifying anomalous regions that deviate from normal across multiple categories. Diffusion-based methods have recently attracted attention due to their excellent performance in this task. However, existing methods concentrate on enhancing the denoising network of the diffusion model by adding more constraints to ensure high consistency in multistep generation and achieve superior reconstruction performance. Additional sampling steps also lead to higher computational costs. We propose a novel multiclass anomaly detection and localization method called TimeNet to address these issues. It is based on the diffusion model of ordinary differential equations and achieves high-quality reconstruction with only one-step generation. We introduce a time-perceptive network to address the consistency and identity shortcut problems that may arise from small sampling steps, which further improves the reconstruction quality. Experiments on the most popular benchmark MVTec-AD dataset demonstrate that our TimeNet competes with the current state-of-the-art methods in terms of accuracy while requiring less computational effort and achieving faster speeds. The high accuracy and real-time performance of TimeNet satisfy the requirements for industrial anomaly detection and localization.

**Keywords:** defect detection; anomaly detection; anomaly localization; diffusion model; denoising network; ordinary differential equations; unsupervised learning; timestep-perceptive

视觉异常检测是一项非常重要的任务, 旨在通过对视觉数据进行处理和分析, 识别出与正常

情况不符的异常情况, 在工业表面缺陷检测<sup>[1-8]</sup>、医学诊断<sup>[9-10]</sup>、视频监控<sup>[11-13]</sup>等领域有着广泛的应用。目前大多数异常检测方法<sup>[14-16]</sup>遵循着一类一模型的范式, 即为不同类别的图像分别训练单独的模型。然而, 随着类别数量的增加, 这种范式

收稿日期: 2024-02-22. 网络出版日期: 2025-01-08.

基金项目: 安徽省重点研究与开发计划项目 (2022107020030).

通信作者: 夏秀山. E-mail: [xiaxiushan@iat.ustc.edu.cn](mailto:xiaxiushan@iat.ustc.edu.cn).

©《智能系统学报》编辑部版权所有

会大大增加训练成本和存储开销<sup>[17]</sup>。于是,近年来研究者<sup>[17-21]</sup>开始逐步探索多类一模型的新范式,也就形成了一个具有挑战性的新任务——多类异常检测和定位。多类异常检测和定位只需要训练一个模型,在多类场景下该模型就能够识别出偏离正常的异常区域,既要检测图像中是否包含异常,又要定位出异常所在区域,还要满足实际应用场景中对高精度和实时性的要求。

由于异常种类繁多,有时甚至未曾预见,具有多样性和复杂性,因此通常使用无监督方法来对正常样本数据在特征空间中的分布进行密度估计,将异常检测视为一种分布外问题。常见的无监督异常检测和定位方法主要分为基于表征和基于重建2大类。基于表征的方法包括基于特征嵌入<sup>[16]</sup>、基于教师学生模型<sup>[22-24]</sup>、基于标准化流<sup>[15,25-27]</sup>等,此类方法假设预训练模型提取的特征表示具有很强的区分性,在特征空间中异常样本和正常样本具有显著的差异。基于重建的方法包括自编码器<sup>[14,28]</sup>、生成对抗网络<sup>[29-30]</sup>和扩散模型(denoising diffusion probabilistic models, DDPMs)<sup>[20-21,31-33]</sup>等,此类方法力求在重建过程中,原有的正常区域尽可能被保留,而可能存在的异常区域被重建或者预测成与之相对应的正常区域。在多类异常检测和定位中,由于基于表征的方法用一个模型直接学习多类的正常样本分布,其边际分布不如单类情形下显著,因此降低了特征表示的区分性。而在基于重建的方法中,扩散模型作为一种新兴的生成模型,在异常检测领域取得了不俗的效果<sup>[31]</sup>。相比于自编码器、变分自编码器和生成对抗网络,其训练过程更加稳定,不仅能在小型数据集上进行训练,还具备强大的重建能力<sup>[34]</sup>。

但是,扩散模型受到反向采样速度慢的限制,通常需要数十步乃至数百步采样过程来生成高质量图像。即使通过去噪扩散隐模型(denoising diffusion implicit models, DDIM)<sup>[35]</sup>等加速采样的手段,目前基于扩散模型的异常检测方法<sup>[20-21,31-33]</sup>仍然需要10步乃至更多步才能实现高精度的异常检测和定位,需要在检测精度和推理速度之间进行权衡。然而,为了满足工业应用场景中多类异常检测和定位的高精度和实时性的要求,不仅需要在维持高精度的同时加快推理速度,还需要使得模型学好每个类别的边际分布,即在保持高质量生成结果的同时尽可能减少扩散模型的生成步数。

事实上,为了使得扩散模型更好地学习到不同类别的边际分布,可以考虑引入条件控制<sup>[36]</sup>的思想,将无条件扩散模型变成条件扩散模型,将

无条件生成转变成条件下的可控生成。这样的扩散模型就能成为多类异常检测和定位的基础重建模型。据此,本文提出了一种基于重建的多类异常检测和定位方法,使用常微分方程(ordinary differential equations, ODE)扩散模型 Rectified Flow<sup>[37-38]</sup>来实现高质量的一步生成,并将类别信息作为条件去控制生成样本的域,以获得更强的边际分布。在网络结构方面,由2个相似的并行 U-Net<sup>[39]</sup>类网络组成:一个是去噪网络,基于 ODE 扩散模型 Rectified Flow,用于高质量的一步生成;一个是时间步感知网络(timestep-perceptive network, TPN),用于辅助修复异常区域。ODE 扩散模型 Rectified Flow 通过学习神经 ODE 模型,在2个经验分布之间直接进行传输,仅需一步反向求解,就可以产生高质量的生成结果。时间步感知网络与去噪网络并行,输入扩散后的带噪图像,输出控制去噪网络的时间步感知特征和时间步长,用于控制生成模型朝着一致性的方向生成,从而尽量避免发生恒等捷径的问题。一致性是指保持生成图像与输入图像的正常区域基本一样,并将可能存在的异常区域重建成跟正常区域一样的模式;恒等捷径指正常区域和异常区域都被模型重建成输入图像本身,即模型丧失了重建异常至正常的能力。

本文的主要贡献总结如下:

1) 本文提出了一种基于 Rectified Flow 的多类异常检测和定位方法,可以通过高质量一步生成实现高性能的多类别异常检测和定位。

2) 本文设计了与去噪网络并行连接的时间步感知网络,在一步生成过程中引导异常区域恢复为正常区域,同时保持其他正常区域的初始状态。

3) 在通用的基准数据集 MVTec-AD<sup>[1]</sup>上进行实验评估,定性和定量结果证明了本文提出的方法在性能上可与目前最优方法相媲美。

## 1 相关工作

### 1.1 多类异常检测和定位

最近几年已有相关工作将目光从最流行的单类转向多类异常检测和定位中来。UniAD<sup>[17]</sup>首次提出一个统一模型来探索其在多类异常检测和定位中的性能表现,构建了一个基于 Transformer 的特征重建算法框架,其中包括逐层查询解码器和近邻掩码注意力模块。OmniAL<sup>[18]</sup>是一种自监督方法,提出了一个统一的卷积神经网络框架,采取参考样本引导的异常合成策略,使用一个自编码器来重建样本区域和图像,还使用一个 U-Net



类网络来进行分割。HVQ-Trans<sup>[19]</sup>也是特征重建方法,基于层次化量化原型的Transformer来解决恒等捷径问题。RAN<sup>[20]</sup>基于扩散模型来去噪和重建,将异常检测任务建模为噪声或异常去除问题,利用扩散模型进行梯度去噪,从异常图像中重构出正常图像,并提供了一种异常检测结果的可解释性分析。DiAD<sup>[21]</sup>则基于潜在扩散模型的方法,构造了一个语义引导网络来连接潜在扩散模型的去噪网络,以保持一致性的语义信息并重建异常,还提供了一个空间感知特征融合模块集成多尺度特征来进一步提升重建能力。

## 1.2 扩散模型

扩散模型<sup>[34,40-41]</sup>由于其强大的生成能力而广泛应用于许多领域,比如文生图、图生图、可控生成、图像编辑、视频生成等。潜在扩散模型(latent diffusion models, LDM)<sup>[42]</sup>在隐空间中执行扩散过程,该空间维度较低,可以降低计算成本。文生图条件控制扩散模型ControlNet<sup>[43]</sup>是一种神经网络架构,用于为预训练的潜在扩散模型增加空间条件控制。就异常检测而言,AnoDDPM<sup>[31]</sup>首次将单纯形噪声引入扩散模型并应用于医学图像异常检测。DiffAD<sup>[32]</sup>提出了基于潜在扩散模型的框架,包括带有噪声条件嵌入和通道插值算法。DDAD<sup>[33]</sup>使用条件控制的机制,引导去噪过程朝着目标图像的方向生成,并引入了特征域适应的方法来微调特征提取网络,从而提升检测精度。这些工作,包括上述的RAN<sup>[20]</sup>和DiAD<sup>[21]</sup>,通常使用DDIM<sup>[35]</sup>的方法来加速多步的反向采样过程,这会引入一个检测精度和推理时间之间的折衷问题。

与现有基于扩散模型的方法不同的是,本文

没有选择常用扩散模型来作为基础重建模型,而是选择更快的ODE扩散模型。近年来在加速扩散模型采样方面,常见的方法有DDIM<sup>[35]</sup>、高阶求解器<sup>[44-45]</sup>、Rectified Flow<sup>[37-38]</sup>、一致性网络<sup>[46]</sup>和蒸馏法<sup>[47]</sup>等。其中Rectified Flow可以视作对每一时间步都平等地进行了蒸馏,对于具有复杂性的真实异常来讲,采取的线性扩散过程便于时间步感知网络更具解释性地一步生成,异常与其期望正常之间的变换缺少的变量正是加噪的前向扩散时间步。Rectified Flow<sup>[37]</sup>引入了一种特殊的Reflow过程,增强了2个经验分布之间的耦合,并将ODE扩散模型的生成过程压缩至一步。Rectified Flow相关研究<sup>[37-38]</sup>表明,该方法确实能够在小型数据集以及大规模文生图任务中实现高质量的一步生成能力。

因此,本文遵循Rectified Flow的基本流程,根据多类异常检测和定位这个任务的特点专门设计了2点改进:一是将Rectified Flow扩展成类别条件控制下的模型,并同样通过Reflow流程增强其一步生成能力;二是独特设计了时间步感知网络,用于感知异常。

## 2 基于ODE扩散模型的多类异常检测和定位算法

本文方法的整体框架如图1所示。在推理阶段,待测图像 $X_i$ 经过扩散过程得到加噪图像 $X_t$ ,将加噪图像同时输入去噪网络和时间步感知网络,分别得到感知时间步 $T$ 和速度场 $v_{k|T}$ ,经过一步生成得到重建图像 $\hat{X}_{1|t}$ ,与待测图像进行比较后得到异常分数图 $M$ 。其中 $\mathcal{E}$ 表示编码器, $\mathcal{D}$ 表示解码器。

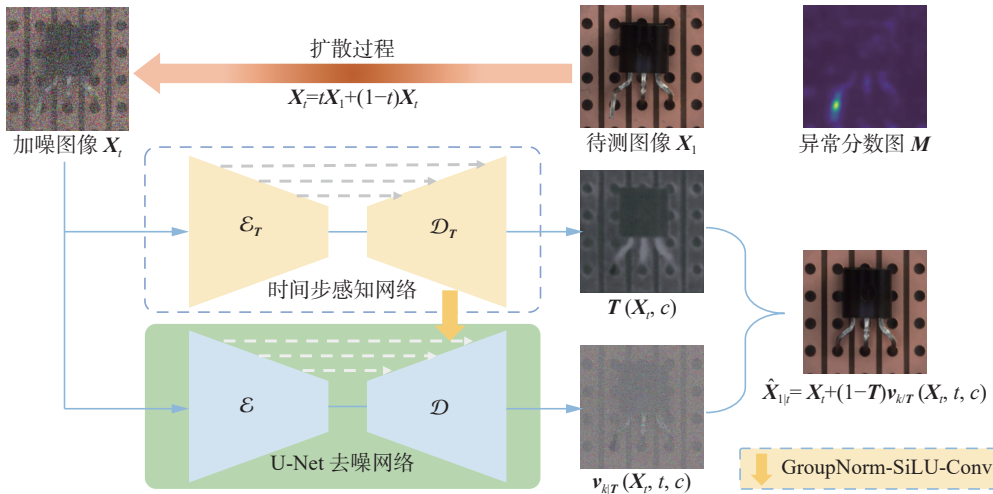


图1 本文方法的整体框架

Fig. 1 Overall architecture of the proposed method

## 2.1 Rectified Flow 原理

### 2.1.1 Rectified Flow

Rectified Flow<sup>[37]</sup> 提供了一种生成式建模的方法, 通过学习神经 ODE 模型可以在 2 个经验分布  $\pi_0 \in \mathbb{R}^d$  和  $\pi_1 \in \mathbb{R}^d$  之间直接传输。给定 2 个观测样本  $X_0 \sim \pi_0$  和  $X_1 \sim \pi_1$ , Rectified Flow 引入了一种简单的线性插值:

$$X_t = tX_1 + (1-t)X_0 \Rightarrow \frac{d}{dt}X_t = X_1 - X_0$$

这是有别于在常见扩散模型中使用的非线性插值过程 ( $X_t = \alpha_t X_0 + \beta_t X_1$ ), 比如 DDPM<sup>[41]</sup>、DDIM<sup>[35]</sup>、LDM<sup>[42]</sup>、Flow ODEs<sup>[48]</sup> 等。注意到  $\frac{d}{dt}X_t$  是对时间步  $t$  独立的, 仅代表从  $X_0$  指向  $X_1$  的线性路径的变化方向 ( $X_1 - X_0$ )。因此, Rectified Flow 设立了一个速度场  $v: \mathbb{R}^d \rightarrow \mathbb{R}^d$ , 通过一个 ODE 模型将从  $\pi_0$  到  $\pi_1$  的轨迹尽可能拉直:

$$\frac{d}{dt}Z_t = v(Z_t, t)$$

这可以通过最小化一个简单的均方差目标函数得到

$$\min_v \int_0^1 E[\|X_1 - X_0 - v(X_t, t)\|^2] dt \quad (1)$$

式中  $v$  可以通过神经网络赋以参数化, 并且式 (1) 可以通过现有的优化算法进行最优化求解。

### 2.1.2 快速生成

使用 Rectified Flow 对  $v$  进行的反向生成过程记为

$$\text{ODE}[v](Z_t) = \hat{Z}_{1/t}(Z_t, t, v) \quad (2)$$

式中  $\text{ODE}[v]$  需要通过数值方法进行模拟求解, 比如 RK45 或者 Euler 法。RK45 指的是龙格库塔法, 可以根据输入的参数自适应决定步长和求解步数。Euler 法则是等步长模拟 ODE 的反向采样过程:

$$Z_{t+\frac{1}{N}} = Z_t + \frac{1}{N}v(Z_t, t), \forall t \in \{0, 1, \dots, N-1\}$$

式中步长为  $1/N$ , 共需要  $N$  步。当步长足够大时, 这个求解过程可以简化成一步:

$$\hat{Z}_{1/t}(Z_t, t) = Z_t + (1-t)v(Z_t, t) \quad (3)$$

因此, 拉直 ODE 轨迹是取得高质量一步生成能力的必要途径。

### 2.1.3 Reflow

Reflow<sup>[37]</sup> 是一种生成微调的迭代程序, 使用  $k$ -Rectified Flow  $v_k$  生成的配对数据 ( $X_0, X_1$ ) 进行微调来拉直  $v_k$  的轨迹, 得到  $v_{k+1}$ 。使得  $v_{k+1}$  具有更直的 ODE 轨迹, 并且分别由  $v_k$  和  $v_{k+1}$  生成的  $X_1$  具有相同的分布。

$$v_{k+1} = \arg \min_v \int_0^1 E[\|(X_1 - X_0) - v(X_t, t)\|^2] dt$$

式中:  $X_1 = \text{ODE}[v_k](X_0)$  是使用  $v_k$  生成的图像,  $X_t = tX_1 + (1-t)X_0$  是前向扩散得到的加噪图像。

### 2.1.4 类别条件控制的 Rectified Flow

像其他条件扩散模型<sup>[36,43]</sup> 一样, 速度场  $v$  可以增加条件信息来控制生成与之相关的图像。本文选择类别作为一种必要的条件来生成指定类别的正常图像。在类别条件控制下, 目标函数变成

$$\min_v \int_0^1 E[\|(X_1 - X_0) - v(X_t, t, c)\|^2] dt \quad (4)$$

式中:  $c$  指与输入图像  $X_1$  相一致的类别条件,  $X_1 = \text{ODE}[v_k](X_0|c)$  是使用 ODE 求解器生成的指定类别为  $c$  的图像,  $X_t = tX_1 + (1-t)X_0$  也是前向扩散得到的加噪图像。

如算法 1 所示, 本文用  $v_0$  表示最开始训练的类别条件控制的 Rectified Flow, 其参数表示为  $\theta_1$ 。后续通过 Reflow 操作微调模型得到  $v_k$ , 其具有高质量一步生成的能力, 被用来作为多类异常检测和定位的基础重建模型。

**算法 1** 训练类别条件控制的  $k$ -Rectified Flow

**输入** 扩散模型去噪网络  $v_0 = v_{\theta_1}$ ; 正常数据集  $D_N$  及其类别标签  $c$ 。

**输出** 扩散模型去噪网络  $v_a$ 。

1) for  $k \leq a$  ( $a$  是 Reflow 次数) do

2) 从  $v_{k-1}$  初始化得到  $v_k$

3) 使用式 (2) 生成耦合数据对 ( $X_0, X_1$ ), 用于 Reflow 程序在  $v_{k-1}$  上训练  $v_k$ , 其中  $X_1$  是使用 RK45 生成的  $X_1 = \text{ODE}[v_{k-1}](X_0 | c)$

4) 通过优化式 (4) 来训练  $v_k$

5) end for

## 2.2 时间步感知网络

正如引言中所讨论的, 基于重建的异常检测和定位方法需要尽可能提高重建质量, 以维持一致性和消除恒等捷径。基于扩散模型的方法<sup>[20-21,31-33]</sup> 的流程首先是将待测图像扩散一定步数得到加噪图像, 然后重建该加噪图像至清晰图像, 通过比较待测图像与重建图像得到表示异常概率大小的异常分数图。然而, 这类方法的实验结果显示扩散步数对重建质量有着不可忽视的影响。因为生成的图像仅仅是从模型所学习到的正常样本分布中采样得到的。扩散步数越多, 生成的图像是正常的, 但越有可能跟输入图像不一致, 比如可能是另一个角度摆放的正常物体。扩散步数越少, 加噪图像越难以完全破坏异常所在区域, 导致生成的图像越有可能仍然保留该异常

区域的信息。

因此,为了解决上述问题,本文设计了一种时间步感知网络(timestep-perceptive network, TPN)来尽可能维持一致性和缓解恒等捷径的问题。TPN是专为多类异常检测与定位设计的模块,灵敏感知异常区域与其期望正常区域之间的差值。本文受到ControlNet<sup>[43]</sup>的一定启发,TPN具备和Rectified Flow<sup>[37]</sup>中使用的去噪网络相似的U-Net<sup>[39]</sup>结构。其输入与去噪网络一样,是扩散后的加噪图像。但不同的是,TPN输入没有扩散时间步 $t$ ,去掉了相关网络层。输出则包括2个部分:一个是输出感知到的扩散时间步 $T$ (与输入图像尺寸一致),从而代替式(3)中的 $t$ 进行生成;另一个部分是输出时间步感知特征跨连接到扩散模型去噪网络对应的尺度中,以达到控制去噪网络输出的目的。预训练好的扩散模型去噪网络参数被冻结,从时间步感知网络中传过来的时间步感知特征需经过一个特征自适应模块,如图1中金色箭头所示。该模块是扩散模型使用的U-Net去噪网络中一个常见的网络模块,由Group-Norm、SiLU和Conv依次连接而得,用于对时间步感知特征进行自适应的跨通道信息交互和平滑。其中,GroupNorm指组归一化(group normalization),它将特征通道数分为若干组,并对每组内的特征进行归一化处理;SiLU(sigmoid linear unit)是一种神经网络的激活函数,由Sigmoid函数乘以其输入计算可得;Conv指 $1 \times 1$ 的卷积层。最后,使用自适应平均池化来对齐2个并行网络的特征分辨率,以传输时间步感知特征。于是,最终的一步生成方式为

$$\hat{X}_{1T}(X_t, t, c) = X_t + (1 - T)v_{kT}(X_t, t, c)$$

式中: $T = T(X_t, c)$ 代表时间步感知网络TPN, $v_{kT}(X_t, t, c)$ 是TPN用来控制输出的预训练扩散模型去噪网络。

为了有效训练时间步感知网络,本文使用与其他自监督重建方法<sup>[10,18,32]</sup>相同的异常合成策略,其人造异常是通过将正常图像和从DTD<sup>[49]</sup>数据集中随机抽取的纹理图像组合得到的。训练时间步感知网络时使用到的重建损失函数是像素级均方误差(mean squared error, MSE)损失和特征级图像块相似度感知(learned perceptual image patch similarity, LPIPS)损失<sup>[50]</sup>之和,其中LPIPS是一种深度特征度量图像相似度的方法,它通过深度学习模型来评估2个图像之间的感知差异。训练流程如算法2所示,前后训练好的ODE扩散模型和

时间感知网络组成了本文提出的模型框架。

## 算法2 基于k-Rectified Flow训练TPN

**输入** 扩散模型去噪网络 $v_k$ ;时间步感知网络 $T = T_{\theta_t}(X_t, c)$ ;正常数据集 $D_N$ 及其类别标签 $c$ 。

**输出** 训练好的时间步感知网络 $T$ 。

1) 初始化 $T$ ;

2) 使用正常数据集 $D_N$ 和纹理图像数据集DTD来合成人造异常样本 $X_a$ ;

3) 训练 $T$ ,损失函数为MSE和LPIPS之和。

## 2.3 异常检测和定位

在推理评估阶段,对于基于重建的方法来说,通过同时在像素级和特征级计算输入图像和重建图像的相似度来得到异常分数图是至关重要的。异常分数图中每个像素的值代表该像素的异常概率,值越大,代表异常可能性越高。为了综合得到异常分数图,本文采用一种加权策略<sup>[33]</sup>,在欧氏距离和余弦相似度之间平衡各自的重要性:

$$F(d_1, d_2, w) = \left( w \frac{\max d_2}{\max d_1} \right) d_1 + d_2 \quad (5)$$

式中: $d_1$ 和 $d_2$ 分别代表计算得到的欧氏距离和余弦相似度, $w$ 是用来控制标准化后的 $d_1$ 重要性的加权系数。标准化后的 $d_1$ 与 $d_2$ 具有相同的上界。

本文将式(5)应用于从预训练网络提取的多尺度特征并求和得到特征级异常分数图 $S_f = \sum_i \sigma_i F(d_{1,i}, d_{2,i}, w_i)$ ,其中 $\sigma_i$ 指上采样至图像尺寸。相似地,结合像素级异常分数图 $S_p = \|X_1 - \hat{X}_{1T}\|_2^2$ ,用 $\lambda$ 控制 $S_p$ 的重要性,得到最终的异常分数图:

$$S = F(S_p, S_f, \lambda)$$

## 3 实验结果与分析

### 3.1 数据集、评价指标和实现细节

#### 3.1.1 MVTec-AD数据集

MVTec-AD<sup>[1]</sup>数据集来自真实世界工业异常检测场景,包含10种物体类和5种纹理类图像共计5354张图像,其中共有73种异常类型,包括划痕、凹陷、污点和其他结构性缺陷等。

#### 3.1.2 评价指标

异常检测和定位需要分别判断整个图像包含异常的概率和每个像素属于异常区域的概率,可视为判断正常和异常的二分类问题。因此,常使用二分类问题的评价指标——受试者工作特征曲线下面积(area under the receiver operating characteristic curve, AUROC),即计算受试者工作特性(receiver operating characteristic, ROC)曲线下的面



积。ROC 曲线的横轴是假正例率 (false positive rate, FPR), 纵轴是真正例率 (true positive rate, TPR)。两者分别定义为

$$\begin{aligned} T_{PR} &= N_{TP}/(N_{TP} + N_{FN}) \\ F_{PR} &= N_{FP}/(N_{TN} + N_{FP}) \end{aligned} \quad (6)$$

式中:  $N_{TP}$  是真正例的数量,  $N_{FP}$  是假正例的数量,  $N_{TN}$  是真反例的数量,  $N_{FN}$  是假反例的数量。

在异常检测中往往将异常样本设置成正样本, 正常样本设置成负样本。根据模型预测得到的异常概率值, 将待测样本进行排序。选取这个排序中间的某个值作为截断点将待测样本分为 2 部分, 高于截断点的判为正例, 反之判为反例。于是  $N_{TP}$  就是判作正例中实际为正样本的数量, 其余同理。选取不同的截断点, 使用式 (6) 计算得到一系列的 TPR 和 FPR, 从而绘制出一条 TPR-FPR 曲线, 即 ROC 曲线。

本文采取 AUROC 分别作为图像级异常检测和像素级异常定位的评价指标, 分别记为 I-AUROC 和 P-AUROC。然而, P-AUROC 对每个像素平等看待, 在计算时侧重于大的异常区域。因此, 本文还对比了异常定位的 PRO(per-region overlap)<sup>[1]</sup> 指标, 它平衡了不同大小的异常区域的重要性, 相比 P-AUROC 更加准确。PRO 对图像中每个真值异常掩码连通域 (即异常区域) 内的异常像素单独考虑, 其表达式为

$$I_{PRO} = \frac{1}{N} \sum_i \sum_k \frac{|P_i \cap C_{i,k}|}{|C_{i,k}|}$$

式中:  $N$  是测试集所有真值异常掩码连通域的数量,  $P_i$  是第  $i$  张待测图像的异常分数图,  $C_{i,k}$  是第  $i$  张待测图像对应的真值异常掩码的第  $k$  个连

通域。

### 3.1.3 实现细节

在 MVTEC-AD 数据集集中的所有图像均被放缩尺寸至 256 像素×256 像素, 且没有使用任何图像增强手段。对于训练类别条件控制的 Rectified Flow, 本文使用跟常见扩散模型一样的 U-Net 网络作为去噪网络。但与默认设定不一样的是, 本文将参数量减少至  $30 \times 10^6$ , 其中基础通道数为 64, 通道数乘积分别为 1、1、2、2、4、4, 注意力模块涉及到的分辨率分别为 8 和 16, 训练过程没有使用指数滑动平均 (exponential moving average, EMA)。如算法 1 和算法 2 所示, 本文选择经过一次 Re-flow 后得到的模型作为基础重建模型, 然后开始训练时间步感知网络, 最终得到用于多类异常检测和定位的全部模型。在推理评估阶段, 初始扩散时间步默认为 0.83, 异常分数图中所有加权系数默认为 7, 并最终通过核为 4 的高斯滤波得到更为平滑的异常分数图。就异常检测而言, 对异常分数图应用 8 轮全局平均池化再取最大值作为图像级异常分数。

### 3.2 定性结果

本文通过一些定性结果来证明本文方法的优异效果。为了尽可能直观地展现模型重建能力, 本文对比了同样的基于重建的方法, 包括非扩散模型的 DRAEM<sup>[14]</sup> 和基于扩散模型的 DiAD<sup>[21]</sup> 方法。图 2~4 给出了本文方法跟这些基于重建的方法的可视化效果对比, 其中“MVTEC-AD”2 列分别指待测图像和真值异常掩码, 其余代表不同的重建方法, 每种方法的 2 列分别指重建图像和对应的异常分数图。

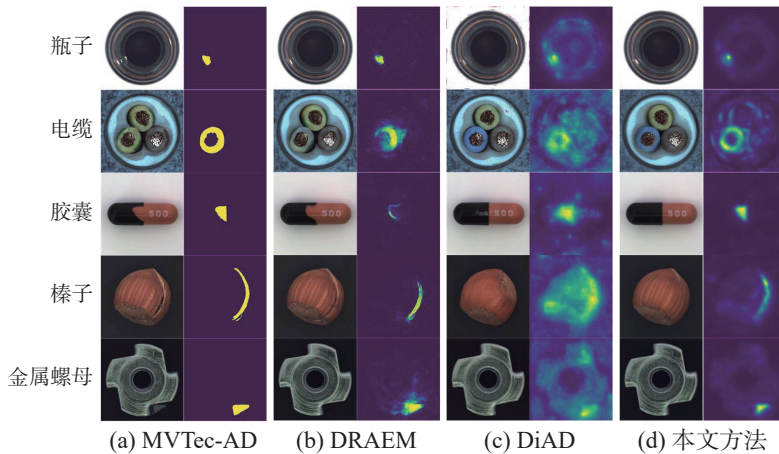


图 2 MVTEC-AD 数据集的可视化

Fig. 2 Visualizations of the MVTEC-AD dataset

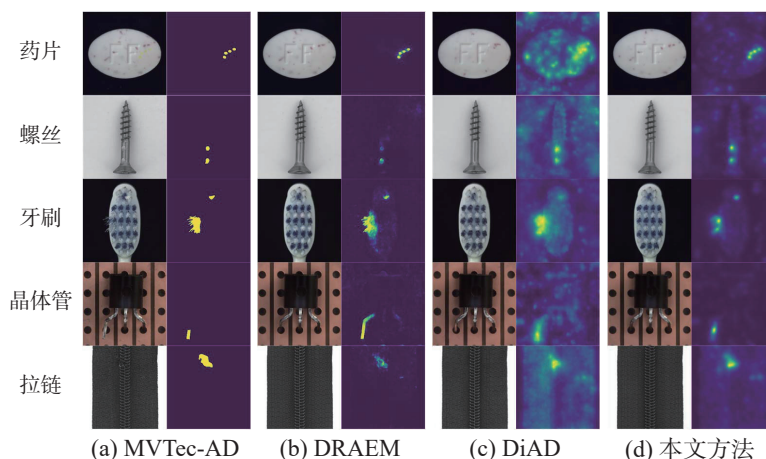


图 3 MVTec-AD 数据集的可视化

Fig. 3 Visualizations of the MVTec-AD dataset

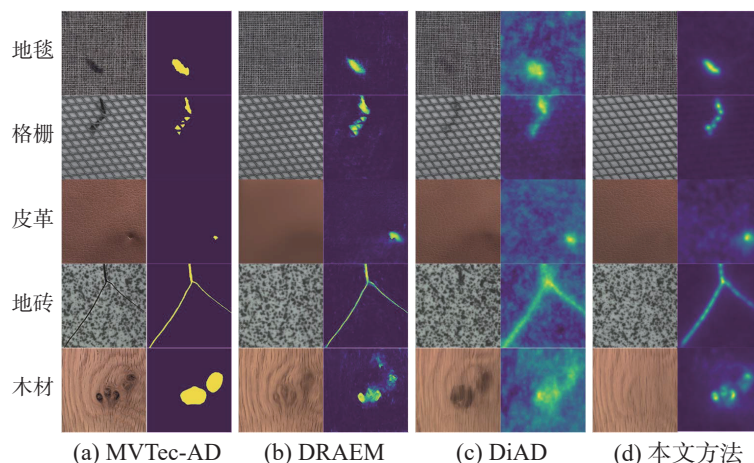


图 4 MVTec-AD 数据集的可视化

Fig. 4 Visualizations of the MVTec-AD dataset

结果显示: DRAEM 的重建效果相对比较模糊, 在纹理类图像 (如图 4 第 3 行皮革) 上模糊效果尤为明显; 此外存在恒等捷径的问题, 模型直接重建缺陷本身, 比如图 2 第 2 行电缆左下方保护层颜色未得到重建, 第 3 行胶囊缺陷处未被修复, 第 4 行榛子破损的外壳未得到复原。DiAD 的 10 步重建质量比 DRAEM 高, 缓解了部分恒等捷径问题, 比如图 3 第 2 行的电缆左下方保护层能够由青绿色重建回正常的蓝色; 但存在一致性问题, 生成的图像与输入图像不匹配, 比如图 2 第 4 行榛子重建后出现 2 个毛状顶部, 且表面纹理细节发生损坏, 第 2 行的电缆, 重建后保护层的形状发生了变化, 图 4 第 4、5 行的纹理类图像重建后正常区域的纹理细节发生了随机的变化。相比较而言, 本文方法仅用 1 步就能生成高质量重建图像, 生成的一致性较高, 还有效缓解了恒等捷径问题, 比如图 2 第 4 行榛子表面破损被修复, 而其余正常区域几乎维持一致; 并且对异常区域

感知的界限更加清晰, 比如图 4 纹理类图像重建图像比较清晰, 除了异常区域被修复外其余正常区域的细小纹理细节也尽可能保持一致。从各类方法的主观结果对比来看, 本文方法总体效果更好。

图 5 给出了不同前向扩散时间步下不同模型的重建结果。第 1 行是指对输入图像加噪不同前向扩散时间步的结果, 随着前向扩散时间步的增大, 给输入图像添加的噪声相应减少, 导致正常的语义信息保留得较多, 而异常区域却破坏得较少。第 2~4 行分别指 Rectified Flow、1-Rectified Flow、本文方法进行一步重建的结果。可以发现, Rectified Flow 和 1-Rectified Flow 均不能有效地重建异常, 将其恢复成正常模样。当噪声水平较高时, 这种纯粹的扩散模型方法倾向于生成正常样本但不一定跟输入图像保持一致。但是, 本文提出的方法具备更优异的重建能力, 能有效缓解上述问题。



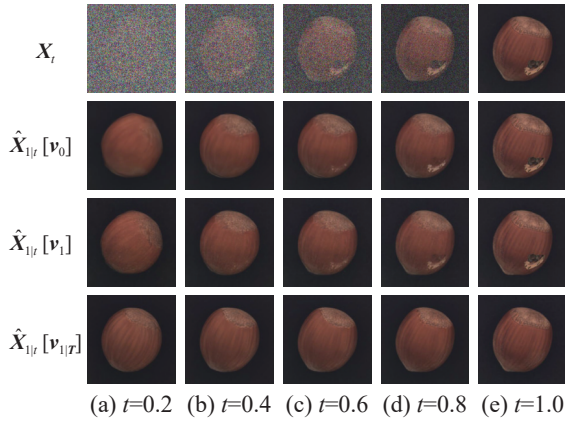


图 5 不同前向扩散时间步的重建结果

Fig. 5 Reconstructions with various forward diffusion timesteps

在实际应用中,往往需要将异常分数图进一步进行阈值分割,将检测出来的异常区域可视化展示出来。图 6 便进一步提供了异常定位的可视化效果图,红色边界表示使用本文方法并以 F1 分数产生的阈值进行异常定位的轮廓。该阈值是由系列查准率和查全率计算得到的最大 F1 分数对应的阈值。

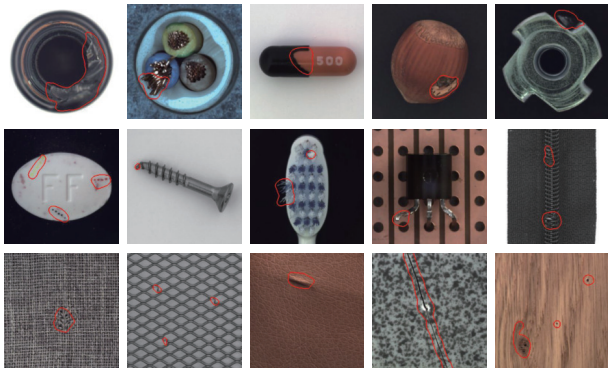


图 6 MVTec-AD 数据集中的例子

Fig. 6 Examples from the MVTec-AD dataset

结果显示,本文方法能够精确地检测并定位到不同样本中的异常,包括不同类别、形状和大小的异常。

### 3.3 定量结果

本文开展了一系列定量实验,在 MVTec-AD 数据集上对比了大量先进方法,从而验证本文方法在 multi-class 异常检测和定位上的客观性能。参与对比的方法中, DRAEM<sup>[14]</sup> 和 PatchCore<sup>[16]</sup> 是单类异常检测和定位中的先进方法,其余是 multi-class 异常检测和定位中的最先进方法,其中 UniAD<sup>[17]</sup>、OmniAL<sup>[18]</sup> 和 HVQ-Trans<sup>[19]</sup> 是非扩散模型法, RAN<sup>[20]</sup> 和 DiAD<sup>[21]</sup> 是基于扩散模型的方法。

表 1 给出了图像级异常检测和像素级异常定位的评价指标结果(括号里分别是 I-AUROC 和 P-AUROC)。粗体为最优结果、下划线为次优结果(下同),“本文方法”指在同一参数下的实验结果,“本文方法\*”指每一类分别调参后的实验结果。整体而言,当对所有类别使用相同的参数设置来进行推理测试时,本文方法取得次优的结果。当针对每一类单独调整参数时,本文方法能够在异常检测上取得最优结果,在异常定位上非常逼近最优结果。其中本文方法在物体类图像取得或接近次优结果,在纹理类图像均取得最优结果。注意到,本文方法仅使用 1 步生成就能打败其他同类型使用 DDIM 10 步生成的基于扩散模型的方法。此外,本文方法的模型参数量是  $60 \times 10^6$ , 同类型 DiAD 参数量多达  $1.4 \times 10^9$ 。本文方法具备模型轻量、检测快、精度高的优点。

表 2 则给出了针对异常定位更准确的 PRO 评价指标下的实验结果,本文方法取得了最优的结果,说明本文方法对异常区域定位得更加准确。

表 1 在 MVTec-AD 数据集上与 SOTA 方法相比的异常检测和定位实验结果

类别	DRAEM <sup>[14]</sup>	PatchCore <sup>[16]</sup>	UniAD <sup>[17]</sup>	OmniAL <sup>[18]</sup>	HVQ-Trans <sup>[19]</sup>	DiAD <sup>[21]</sup>	本文方法	本文方法*
Bottle	(97.5,87.6)	<b>(100.0,97.4)</b>	(99.7,98.1)	<b>(100.0,99.2)</b>	<b>(100.0,98.3)</b>	(99.7,98.4)	<b>(100.0,98.8)</b>	<b>(100.0,98.9)</b>
Cable	(57.8,71.3)	(95.3,93.6)	(95.2,97.3)	<u>(98.2,97.3)</u>	<b>(99.0,98.1)</b>	(94.8,96.8)	(91.8, <u>97.4</u> )	(92.6, <u>97.4</u> )
Capsule	(65.3,50.5)	<b>(96.8,98.0)</b>	(86.9, <u>98.5</u> )	(95.2,96.9)	<u>(95.4,98.8)</u>	(89.0,97.1)	(91.6,96.7)	(93.7,97.3)
Hazelnut	(93.7,96.9)	(99.3,97.6)	<u>(99.8,98.1)</u>	(98.7, <b>99.4</b> )	<b>(100.0,98.8)</b>	(99.5,98.3)	(98.8,98.6)	(98.8, <u>98.8</u> )
Metal Nut	(72.8,62.2)	(99.1,96.3)	(99.2,94.8)	<b>(99.9,99.4)</b>	<b>(99.9,96.3)</b>	(99.1, <u>97.3</u> )	(99.1,96.6)	<u>(99.7,96.6)</u>
物体类 Pill	(82.2,94.4)	(86.4,90.8)	(93.7,95.0)	(95.6, <b>98.4</b> )	(95.8,97.1)	(95.7,95.7)	<u>(96.2,97.5)</u>	<b>(97.4,97.7)</b>
Screw	(92.0,95.5)	(94.2,98.9)	(87.5,98.3)	<b>(99.0,99.3)</b>	<u>(95.6,98.9)</u>	(90.7,97.9)	(90.9,99.2)	(94.1, <b>99.5</b> )
Toothbrush	(90.6,97.7)	<b>(100.0,98.8)</b>	(94.2,98.4)	(99.2, <u>99.1</u> )	(93.6,98.6)	<u>(99.7,99.0)</u>	<b>(100.0,99.2)</b>	<b>(100.0,99.2)</b>
Transistor	(74.8,64.5)	(98.9,92.3)	<u>(99.8,97.9)</u>	(97.2, <b>98.9</b> )	(99.7, <u>97.9</u> )	<u>(99.8,95.1)</u>	(99.0,94.0)	<b>(100.0,94.2)</b>
Zipper	(98.8,98.3)	(97.1,95.7)	(95.8,96.8)	(88.0,98.0)	(97.9,97.5)	(95.1,96.2)	<b>(99.0,98.3)</b>	<b>(99.0,98.5)</b>
平均值	(82.6,81.9)	(96.7,95.9)	(95.2,97.3)	(97.1, <b>98.6</b> )	<b>(97.7,98.0)</b>	(96.3,97.2)	(96.6,97.6)	(97.5,97.8)

续表1

类别	DRAEM <sup>[14]</sup>	PatchCore <sup>[16]</sup>	UniAD <sup>[17]</sup>	OmniAL <sup>[18]</sup>	HVQ-Trans <sup>[19]</sup>	DiAD <sup>[21]</sup>	本文方法	本文方法*
纹理类	Carpet	(98.0,98.6)	(97.0,98.1)	(99.8,98.5)	(99.6,99.0)	(99.9,98.7)	(99.4,98.6)	(97.9,99.2)
	Grid	(99.3,98.7)	(91.4,98.4)	(98.2,96.5)	(100.0,99.4)	(97.0,97.0)	(98.5,96.6)	(100.0,99.4)
	Leather	(98.7,97.3)	(100.0,99.2)	(100.0,98.8)	(93.8,93.3)	(100.0,98.8)	(99.8,98.8)	(99.9,99.3)
	Tile	(99.8,98.0)	(96.0,90.3)	(99.3,91.8)	(93.2,97.4)	(99.2,92.2)	(96.8,92.4)	(99.9,97.9)
	Wood	(99.8,96.0)	(93.8,90.8)	(98.6,93.2)	(100.0,99.5)	(97.2,92.4)	(99.7,93.3)	(98.6,96.5)
	平均值	(99.1,97.7)	(95.6,95.4)	(99.2,95.8)	(97.3,97.7)	(98.7,95.8)	(98.8,95.9)	(99.3,98.5)
平均值	(88.1,87.2)	(96.4,95.7)	(96.5,96.8)	(97.2,98.3)	(98.0,97.3)	(97.2,96.8)	(97.5,97.9)	(98.2,98.1)

表2 PRO评价指标的实验结果  
Table 2 PRO metric on MVTec-AD

%

评价指标	基于非扩散模型			基于扩散模型		
	DRAEM <sup>[14]</sup>	PatchCore <sup>[16]</sup>	UniAD <sup>[17]</sup>	RAN <sup>[20]</sup>	DiAD <sup>[21]</sup>	本文方法
PRO	71.1	93.5	90.0	93.0	90.7	93.9

### 3.4 消融实验

本小节设置了一系列消融实验来验证本文方法模块和参数等的有效性。

#### 3.4.1 时间步感知网络的影响

如表3所示,本文验证了时间步感知网络的重要性,其中初始扩散时间步均设为0.80。数字1、3、5、10指在反向采样过程中使用的采样步数。结果显示出了3点现象:1)只需一步采样就能达到比多步采样更高的结果。不同采样步数的Rectified Flow结果相近,但随着采样步数的增多,评价指标反而轻微下降。每一次采样会重新推理一次模型,当采样步数增加时,模型推理堆积的误差

会随之增大,导致图像重建的一致性降低,因此评价指标略微下降。2)相同采样步数下1-Rectified Flow比Rectified Flow的评价指标高,这是由于Rectified Flow经过一次Reflow流程后,拉直了ODE扩散模型的轨迹,从而提升了模型的一步生成能力。3)基于1-Rectified Flow增加时间步感知网络,显著提升了各项指标,其中I-AUROC提升了5.6%,P-AUROC提升了2.3%,PRO提升了6.5%。定量实验结果表明,Rectified Flow作为ODE扩散模型,其一步生成和多步生成的效果相近,与预期结果基本一致。本文方法提出的时间步感知模块显著地提升了异常检测和定位的性能。

表3 时间步感知网络的消融实验  
Table 3 Ablation studies on TPN

%

评价指标	Rectified Flow				1-Rectified Flow				本文方法
	1	3	5	10	1	3	5	10	1
I-AUROC	90.0	89.4	89.4	88.6	92.2	92.1	92.0	91.9	97.4
P-AUROC	94.6	95.1	95.1	95.3	95.7	95.7	95.7	95.7	97.9
PRO	86.4	86.3	86.3	85.9	88.1	88.1	88.1	88.0	93.8

#### 3.4.2 预训练特征提取器的影响

如表4所示,使用不同预训练特征提取器,异常分数图取得的结果不同。ResNet类<sup>[51-52]</sup>预训练

模型相对VGG类<sup>[53]</sup>和EfficientNet类<sup>[54]</sup>取得更好结果。因此,在其他实验中,本文选择效果最好的WideResNet101作为默认特征提取器。

表4 不同预训练特征提取器的消融实验  
Table 4 Ablation studies on different feature extractors

%

评价指标	VGG <sup>[53]</sup>		ResNet <sup>[51]</sup>			WideResNet <sup>[52]</sup>		EfficientNet <sup>[54]</sup>		
	16	19	18	34	50	101	50	101	b0	b2
I-AUROC	95.2	95.7	97.5	97.6	97.6	97.2	97.2	97.5	91.0	89.6
P-AUROC	96.8	96.5	97.7	97.8	97.7	97.1	97.5	97.9	93.5	93.3
PRO	91.7	91.6	93.5	93.8	93.6	91.9	93.3	93.9	85.4	84.2

### 3.4.3 多尺度特征层的影响

本文在预训练模型 WideResNet101 下使用不同尺度特征层来计算异常分数图。如表 5 所示,最高层特征  $f_5$  影响非常小,说明其对表面异常的区分度信息较少;最低层特征  $f_1$  会导致指标下滑,这可能是因为分辨率 128 像素 $\times$ 128 像素的特征含有冗余信息,干扰了距离度量。取得最佳效果的是特征层  $f_2$ 、 $f_3$ 、 $f_4$  多尺度结合起来,其对应的特征分辨率分别是  $64 \times 64$ 、 $32 \times 32$ 、 $16 \times 16$ 。

表 5 多尺度特征层的消融实验

Table 5 Ablation studies on multi-scale feature layers %

$f_1$	$f_2$	$f_3$	$f_4$	$f_5$	I-AUROC	P-AUROC	PRO
√	√	√	√	√	97.0	97.0	92.7
√	√	√	√		97.0	97.0	93.7
	√	√	√	√	<b>97.5</b>	<b>97.9</b>	<b>93.9</b>
	√	√	√		<b>97.5</b>	<b>97.9</b>	<b>93.9</b>
	√	√			97.0	97.0	92.7
		√	√		97.2	<b>97.9</b>	93.7

### 3.4.4 前向扩散时间步的影响

初始前向扩散时间步影响着给输入图像加噪的水平。定量结果如图 7 所示,横轴是前向扩散时间步,是  $[0,1]$  区间内的无量纲量,越小代表噪声水平越高,纵轴是不同评价指标 (I-AUROC、P-AUROC 和 PRO)。随着前向扩散时间步逐步增大,各项评价指标呈现先上升后下降的趋势,最优结果位于 80%~85%。

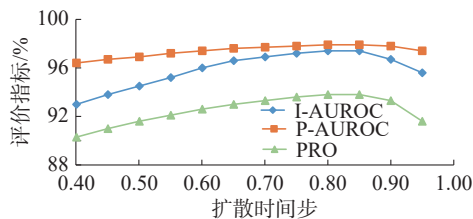


图 7 前向扩散时间步的消融实验

Fig. 7 Ablation on forward diffusion timesteps

图 5 中定性结果也可可视化地给出了该趋势。前向扩散时间步越大,加噪越少,正常语义信息保留得越多,异常区域被破坏的程度越小,导致发生模型恒等捷径问题。反之前向扩散时间步越小,加噪越多,尽管异常区域被破坏的程度变大,但正常语义信息保留变少,导致降低了图像重建一致性。因此选择适当的初始前向扩散时间步有助于取得最优结果。

## 4 结束语

本文提出了一种 ODE 扩散模型方法,用于多

类异常检测和定位。基于 ODE 扩散模型 Rectified Flow, 本文设计了一种专门用于异常检测和定位的时间步感知网络,目的是通过感知时间步控制扩散模型去噪网络的中间时间步嵌入特征,从而取得高质量的一步生成结果,在维持一致性和缓解恒等捷径方面获得优异效果。在通用的工业异常检测数据集 MVTEC-AD 上开展的一系列实验证明了本文方法具备与目前最优方法相媲美甚至更好的性能,尤其是显著优于同类型基于扩散模型的方法,不仅更快而且更好。随着研究的不断深入,未来可在本文方法基础上探索持续学习方法,以面对实际应用中出现新的图像类别和新的异常模式的场景。

## 参考文献:

- [1] BERGMANN P, FAUSER M, SATTLEGGER D, et al. MVTEC AD—a comprehensive real-world dataset for unsupervised anomaly detection[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE, 2019: 9584–9592.
- [2] 吕承侃, 沈飞, 张正涛, 等. 图像异常检测研究现状综述[J]. 自动化学报, 2022, 48(6): 1402–1428.  
LYU Chengkan, SHEN Fei, ZHANG Zhengtao, et al. Review of image anomaly detection[J]. Acta automatica sinica, 2022, 48(6): 1402–1428.
- [3] 陶显, 侯伟, 徐德. 基于深度学习的表面缺陷检测方法综述[J]. 自动化学报, 2021, 47(5): 1017–1034.  
TAO Xian, HOU Wei, XU De. A survey of surface defect detection methods based on deep learning[J]. Acta automatica sinica, 2021, 47(5): 1017–1034.
- [4] 伍麟, 郝鸿宇, 宋友. 基于计算机视觉的工业金属表面缺陷检测综述[J]. 自动化学报, 2024, 50(7): 1261–1283.  
WU Lin, HAO Hongyu, SONG You. A review of metal surface defect detection based on computer vision[J]. Acta automatica sinica, 2024, 50(7): 1261–1283.
- [5] 黄健, 郑春厚, 章军, 等. 基于小样本度量迁移学习的表面缺陷检测[J]. 模式识别与人工智能, 2021, 34(5): 407–414.  
HUANG Jian, ZHENG Chunhou, ZHANG Jun, et al. Few-shot metric transfer learning network for surface defect detection[J]. Pattern recognition and artificial intelligence, 2021, 34(5): 407–414.
- [6] 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.  
XING Peng, JIANG Xin, PAN Yonghua, et al. Feature constrained restricted distillation learning for visual anomaly detection[J]. Journal of software, 2023, 34(9): 4378–4391.



- [7] 王延春秋, 葛泉波, 刘华平. 石板材表面缺陷检测的无监督学习方法[J]. 智能系统学报, 2023, 18(6): 1344–1351.  
WANG Yanchunqiu, GE Quanbo, LIU Huaping. Unsupervised learning method for surface defect detection of slate materials[J]. *CAAI transactions on intelligent systems*, 2023, 18(6): 1344–1351.
- [8] 孙博言, 王洪元, 刘乾, 等. 基于多尺度和注意力机制的混合监督金属表面缺陷检测[J]. 智能系统学报, 2023, 18(4): 886–893.  
SUN Boyan, WANG Hongyuan, LIU Qian, et al. Hybrid supervised metal surface defect detection based on multi-scale and attention[J]. *CAAI transactions on intelligent systems*, 2023, 18(4): 886–893.
- [9] FERNANDO T, GAMMULLE H, DENMAN S, et al. Deep learning for medical anomaly detection—a survey[J]. *ACM computing surveys*, 2022, 54(7): 1–37.
- [10] 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(自然科学版), 2024, 49(6): 851–860.  
HU Xianyao, JIN Congming. Medical image anomaly detection based on diffusion ordinary differential equations[J]. *Journal of Zhejiang Sci-Tech University (natural sciences edition)*, 2024, 49(6): 851–860.
- [11] NAYAK R, PATI U C, DAS S K. A comprehensive review on deep learning-based methods for video anomaly detection[J]. *Image and vision computing*, 2021, 106: 104078.
- [12] 胡正平, 赵梦瑶, 辛丙一. 结合全局与局部视频表示的视频异常检测算法[J]. 模式识别与人工智能, 2020, 33(2): 133–140.  
HU Zhengping, ZHAO Mengyao, XIN Bingyi. Video anomaly detection algorithm combining global and local video representation[J]. *Pattern recognition and artificial intelligence*, 2020, 33(2): 133–140.
- [13] 黄超, 胡志军, 徐勇, 等. 基于视觉的车辆异常行为检测综述[J]. 模式识别与人工智能, 2020, 33(3): 234–248.  
HUANG Chao, HU Zhijun, XU Yong, et al. Vision-based abnormal vehicle behavior detection: a survey[J]. *Pattern recognition and artificial intelligence*, 2020, 33(3): 234–248.
- [14] ZAVRTANIK V, KRISTAN M, SKOČAJ D. DRÆM—A discriminatively trained reconstruction embedding for surface anomaly detection[C]//2021 IEEE/CVF International Conference on Computer Vision. Montreal: IEEE, 2021: 8310–8319.
- [15] GUDOVSKIY D, ISHIZAKA S, KOZUKA K. CFLOW-AD: real-time unsupervised anomaly detection with localization *via* conditional normalizing flows[C]//2022 IEEE/CVF Winter Conference on Applications of Computer Vision. Waikoloa: IEEE, 2022: 1819–1828.
- [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. New Orleans: IEEE, 2022: 14298–14308.
- [17] YOU Zhiyuan, CUI Lei, SHEN Yujun, et al. A unified model for multi-class anomaly detection[C]//Proceedings of the 36th International Conference on Neural Information Processing Systems. New Orleans: Curran Associates Inc., 2022: 4571–4584.
- [18] ZHAO Ying. OmniAL: a unified CNN framework for unsupervised anomaly localization[C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver: IEEE, 2023: 3924–3933.
- [19] LU Ruiying, WU Yujie, TIAN Long, et al. Hierarchical vector quantized transformer for multi-class unsupervised anomaly detection[C]//Proceedings of the 37th International Conference on Neural Information Processing Systems. New Orleans: Curran Associates Inc., 2023: 8487–8500.
- [20] LU Fanbin, YAO Xufeng, FU C W, et al. Removing anomalies as noises for industrial defect localization[C]//2023 IEEE/CVF International Conference on Computer Vision. Paris: IEEE, 2023: 16120–16129.
- [21] HE Haoyang, ZHANG Jiangning, CHEN Hongxu, et al. DiAD: a diffusion-based framework for multi-class anomaly detection[EB/OL]. (2023–12–11)[2024–02–22]. <https://arxiv.org/abs/2312.06607>.
- [22] WANG Guodong, HAN Shumin, DING Errui, et al. Student-teacher feature pyramid matching for anomaly detection[EB/OL]. (2021–03–07)[2024–02–22]. <https://arxiv.org/abs/2103.04257>.
- [23] DENG Hanqiu, LI Xingyu. Anomaly detection via reverse distillation from one-class embedding[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans: IEEE, 2022: 9727–9736.
- [24] TIEN T D, NGUYEN A T, TRAN N H, et al. Revisiting reverse distillation for anomaly detection[C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver: IEEE, 2023: 24511–24520.
- [25] 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. Waikoloa: IEEE, 2021: 1906–1915.
- [26] LEI Jiarui, HU Xiaobo, WANG Yue, et al. PyramidFlow: high-resolution defect contrastive localization using pyramid normalizing flow[C]//2023 IEEE/CVF Conference

- on Computer Vision and Pattern Recognition. Vancouver: IEEE, 2023: 14143–14152.
- [27] 毛国君, 吴星臻, 邢树礼. 基于多尺度流模型的视觉异常检测研究[J]. 自动化学报, 2024, 50(3): 640–648.
- MAO Guojun, WU Xingzhen, XING Shuli. Research on visual anomaly detection based on multi-scale normalizing flow[J]. Acta automatica sinica, 2024, 50(3): 640–648.
- [28] GONG Dong, LIU Lingqiao, LE V, et al. Memorizing normality to detect anomaly: memory-augmented deep autoencoder for unsupervised anomaly detection[C]//2019 IEEE/CVF International Conference on Computer Vision. Seoul: IEEE, 2019: 1705–1714.
- [29] AKCAY S, ATAPOUR-ABARGHOU EI A, BRECKON T P. GANomaly: semi-supervised anomaly detection via adversarial training[C]//14th Asian Conference on Computer Vision. Perth: Springer International Publishing, 2019: 622–637.
- [30] PERERA P, NALLAPATI R, XIANG Bing. OCGAN: one-class novelty detection using GANs with constrained latent representations[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE, 2019: 2893–2901.
- [31] 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. New Orleans: IEEE, 2022: 649–655.
- [32] ZHANG Xinyi, LI Naiqi, LI Jiawei, et al. Unsupervised surface anomaly detection with diffusion probabilistic model[C]//2023 IEEE/CVF International Conference on Computer Vision. Paris: IEEE, 2023: 6759–6768.
- [33] MOUSAKHAN A, BROX T, TAYYUB J. Anomaly detection with conditioned denoising diffusion models [EB/OL]. (2023–05–25)[2024–02–22]. <https://arxiv.org/abs/2305.15956>.
- [34] DHARIWAL P, NICHOL A. Diffusion models beat gans on image synthesis[C]//Proceedings of the 35th International Conference on Neural Information Processing Systems. Vancouver: Curran Associates Inc., 2021, 34: 8780–8794.
- [35] SONG Jiaming, MENG Chenlin, Ermon S. Denoising diffusion implicit models[C]//International Conference on Learning Representations. Vienna: OpenReview.net, 2021: 1–20.
- [36] NICHOL A, DHARIWAL P. Improved denoising diffusion probabilistic models[C]//International Conference on Machine Learning. Virtual: PMLR, 2021: 8162–8171.
- [37] LIU Xingchao, GONG Chengyue, LIU Qiang. Flow straight and fast: learning to generate and transfer data with rectified flow[C]//International Conference on Learning Representations. Kigali: OpenReview. net, 2023: 1–33.
- [38] LIU Xingchao, ZHANG Xiwen, MA Jianzhu, et al. InstaFlow: one step is enough for high-quality diffusion-based text-to-image generation[EB/OL]. (2023–09–12)[2024–02–22]. <https://arxiv.org/abs/2309.06380>.
- [39] RONNEBERGER O, FISCHER P, BROX T. U-net: convolutional networks for biomedical image segmentation [C]//Medical Image Computing and Computer-Assisted Intervention. Munich: Springer International Publishing, 2015: 234–241.
- [40] SOHL-DICKSTEIN J, WEISS E A, MAHESWARA NATHAN N, et al. Deep unsupervised learning using nonequilibrium thermodynamics[C]//International Conference on Machine Learning. Lille: PMLR, 2015: 2256–2265.
- [41] HO J, JAIN A, ABBEEL P. Denoising diffusion probabilistic models[C]//Proceedings of the 34th International Conference on Neural Information Processing Systems. Virtual: Curran Associates Inc., 2020, 33: 6840–6851.
- [42] 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. New Orleans: IEEE, 2022: 10674–10685.
- [43] ZHANG Lyumin, RAO Anyi, AGRAWALA M. Adding conditional control to text-to-image diffusion models[C]//2023 IEEE/CVF International Conference on Computer Vision. Paris: IEEE, 2023: 3813–3824.
- [44] LU Cheng, ZHOU Yuhao, BAO Fan, et al. Dpm-solver: a fast ode solver for diffusion probabilistic model sampling in around 10 steps[C]//Proceedings of the 36th International Conference on Neural Information Processing Systems. New Orleans: Curran Associates Inc., 2022, 35: 5775–5787.
- [45] ZHOU Zhenyu, CHEN Defang, WANG Can, et al. Fast ODE-based sampling for diffusion models in around 5 steps[EB/OL]. (2023–11–30)[2024–02–22]. <https://arxiv.org/abs/2312.00094>.
- [46] SONG Yang, DHARIWAL P, CHEN M, et al. Consistency models[EB/OL]. (2023–03–02)[2024–02–22]. <https://arxiv.org/abs/2303.01469>.
- [47] YIN Tianwei, GHARBI M, ZHANG R, et al. One-step diffusion with distribution matching distillation[EB/OL]. (2023–11–30)[2024–02–22]. <https://arxiv.org/abs/2311.18828>.
- [48] SONG Yang, SOHL-DICKSTEIN J, KINGMA D P, et al.

- Score-based generative modeling through stochastic differential equations[C]//International Conference on Learning Representations. Vienna: OpenReview.net, 2021: 1–36.
- [49] CIMPOI M, MAJI S, KOKKINOS I, et al. Describing textures in the wild[C]//2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus: IEEE, 2014: 3606–3613.
- [50] ZHANG R, ISOLA P, EFROS A A, et al. The unreasonable effectiveness of deep features as a perceptual metric [C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018: 586–595.
- [51] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016: 770–778.
- [52] ZAGORUYKO S, KOMODAKIS N. Wide residual networks[EB/OL]. (2016–05–03)[2024–02–22]. <https://arxiv.org/abs/1605.07146>.
- [53] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[C]//International Conference on Learning Representations. San Diego: OpenReview.net, 2015: 1–14.
- [54] TAN Mingxing, LE Q. EfficientNet: rethinking model

scaling for convolutional neural networks[C]//Proceedings of the 36th International Conference on Machine Learning. Los Angeles: PMLR, 2019: 6105–6114.

#### 作者简介:



蒋世杰, 硕士研究生, 主要研究方向为计算机视觉、异常检测。E-mail: [jiangshijie@mail.ustc.edu.cn](mailto:jiangshijie@mail.ustc.edu.cn)。



夏秀山, 副研究员, 主要研究方向为数字图像处理、计算机视觉、人工智能。主持、参与国家和省部级科研项目 10 余项, 发表学术论文 10 余篇。E-mail: [xiaxiushan@iat.ustc.edu.cn](mailto:xiaxiushan@iat.ustc.edu.cn)。



翟伟, 博士后研究员, 主要研究方向为计算机视觉、具身智能、多模态分析。主持国家自然科学基金青年项目 1 项, 发表学术论文 40 余篇。E-mail: [wzhai056@ustc.edu.cn](mailto:wzhai056@ustc.edu.cn)。