

# 点云与图像融合的无监督异常检测算法研究

谢宏兴<sup>1,2</sup>, 林珊玲<sup>1,2</sup>, 林志贤<sup>1,2</sup>, 郭太良<sup>2</sup>, 林坚普<sup>1,2</sup>, 吕珊红<sup>1,2</sup>

(1. 福州大学 先进制造学院, 福建 泉州 362251; 2. 中国福建光电信息科学与技术创新实验室, 福建 福州 350116)

**摘要:**针对多模态工业检测中高维特征之间存在干扰,导致检测率不理想的问题,提出了一种基于标准化流的多模态工业异常检测方法。首先提取图像的3D点云的深度信息将其作为第4个通道添加到RGB图像中,生成融合后的RGBD图像,然后使用预训练的特征提取网络提取融合后的图像特征,最后使用特征训练得到一个用于异常检测的标准化流模型。实验结果表明,异常检测模型在MVTec 3D-AD数据集的平均Pixel AUROC达到95.8%,平均AUPRO达到86.2%。相较于其他模型分别提升了2.6%和9.1%。

**关键词:**工业异常检测;多模态;点云;标准化流

**中图分类号:**TP391;TN247 **文献标识码:**A **DOI:**10.3969/j.issn.1001-5078.2024.10.020

## Research on unsupervised anomaly detection algorithm for point cloud and image fusion

XIE Hong-xing<sup>1,2</sup>, LIN Shan-ling<sup>1,2</sup>, LIN Zhi-xian<sup>1,2</sup>, GUO Tai-liang<sup>2</sup>, LIN Jian-pu<sup>1,2</sup>, LV Shan-hong<sup>1,2</sup>

(1. School of Advanced Manufacturing, Fuzhou University, Quanzhou 362251, China;

2. Fujian Science and Technology Innovation Laboratory for Photoelectric Information, Fuzhou 350116, China)

**Abstract:** A multimodal industrial anomaly detection method based on normalizing flow is proposed to address the issue of interference between high-dimensional features in multimodal industrial detection, resulting in unsatisfactory detection rates. Firstly, the depth information of the 3D point cloud of the image is extracted and added to the RGB image as the fourth channel to generate the fused RGBD image. Then, the fused image features are extracted using a pre-trained feature extraction network. Finally, a normalizing flow model for anomaly detection is obtained using feature training. The experimental results show that the anomaly detection model achieves an average Pixel AUROC of 95.8% and an average AUPRO of 86.2% on the MVTec 3D-AD dataset, which is an improvement of 2.6% and 9.1%, respectively, compared to other models.

**Keywords:** industrial anomaly detection; multimodal; point cloud; normalizing flow

### 1 引言

工业异常检测的主要目的是发现产品的异常部分,这关系到产品的质量是否符合要求,所以在大规模工业生产的产品质量检验中起着至关重要的作用。目前的工业异常检测方法大多为无监督方法,这是由数据的分布特点决定的,并且异常样本的异

常原因也各不相同,无法穷举出异常的种类。现有的大多数工业异常检测方法都是基于RGB图像的,但是对于例如凹陷和孔洞等缺陷异常,无法仅通过RGB图像的二维信息获取确定,所以三维形状信息对于正确检测异常非常重要,仅从RGB图像是很难识别到缺陷异常的。在MVTec-3D AD数据集<sup>[1]</sup>之

**基金项目:**国家重点研发资助项目(No. 2021YFB3600603);福建省自然科学基金资助项目(No. 2020J01468)资助。

**作者简介:**谢宏兴(1997-),男,硕士研究生,主要研究方向为计算机视觉和异常检测。E-mail:1838040746@qq.com

**通讯作者:**林志贤,男,博士,教授,博导,主要研究方向为光电信息技术、新型显示技术、嵌入式系统设计、图像处理和平板显示驱动技术。E-mail:lzx2005000@163.com

**收稿日期:**2023-12-11; **修订日期:**2024-01-09

前相关的3D数据集仅有BRATS与ATLAS这两个数据集,这两个数据集都是通过堆叠多个灰度图像以形成密集的体素网格<sup>[2]</sup>(Dense Voxel Grid)来提供3D信息,与MVTec 3D-AD集中描述物体几何表面数据有着本质的不同。

针对以上问题,本文提出了一种新的基于标准化流的多模态工业异常检测方法(Anomaly Detection Normalizing Flows,ADNF)。不同于现有基于特征的异常检测方法,ADNF提取了点云的深度信息,融合到对应的RGB图像中,提取融合后的图像的特征并用于训练标准化流模型<sup>[3]</sup>,最后使用模型进行异常检测,这可以使得多模态特征之间的干扰大大减少,从而更好地利用不同模态的互补信息。然后使用经过修改后的预训练特征提取网络提取融合后的图像特征,该特征不仅具有原有的RGB图像的信息,还具有原有的RGB图像所不具备的深度空间信息,所以可以实现二者的信息互补,从而使模型具备有更加准确的异常检测能力。

在本文中,采用的经过预训练的图像特征提取网络(wide\_resnet50\_2)进行适当修改后提取融合后的图像特征。

ADNF的主要工作如下:

(a)提出了一种基于标准化流的多模态工业异常检测方法(ADNF),该方法在MVTec-3D AD数据集上的检测和分割精度优于其他多模态方法。

(b)提出使用点云的空间深度信息辅助图像的异常检测方法,将点云和图像二者在异常检测中的优势互补。

(c)提出将多模态特征用于训练标准化流模型,并将该标准化流模型用来对测试评估样本的异常检测,以实现更加准确和精确的异常检出与定位。

## 2 相关研究

现有的大部分异常检测方法的核心思想是找到正常数据与异常数据之间不同之处的表示方法。比较主流的方法可以分为三类,基于重建的方法<sup>[4]</sup>、基于知识蒸馏的方法<sup>[5]</sup>和基于特征表示的方法<sup>[6]</sup>。Vitjan Zavrtanika等<sup>[7]</sup>将输入图像分割成多个块(patch),并对每个块使用图像修复技术填充缺失区域生成修复后的图像,然后计算输入图像块与对应修复后图像块之间的重构误差,最后通过判断重构误差是否高于阈值来确定该图像块是否包含异常。

重建法认为这些生成模型仅会对正常样本进行很好的重建,然而对于训练过程中没有出现过异常样本中的异常部分不能重建,故对于异常样本的生成跟原有异常样本的差异会变得很大,从而进行异常判别。但是由于某些较小的异常,重建模型具有一定的鲁棒性,并且对于点云和图像两种模态结合的输入,重建模型的训练很难找到合适的方法重建输入目标。基于知识蒸馏的方法上,Guodong Wang等<sup>[5]</sup>提出了一种Teacher-Student semi-supervised学习框架。Teacher模型从正常样本中学习特征表达,Student模型在Teacher指导下学习区分正常和异常样本。由于知识蒸馏的过程需要针对数据类型及特点进行设计,否则学生模型的检测能力不能很好得从复杂的教师模型中学到,并且知识蒸馏过程会增加训练时间和资源消耗。基于特征表示方法中,Jiawei Yu等<sup>[8]</sup>构建了一个2D条件正常化流模型,包含编码器和基于重参数化的2D流,可以对图像进行像素级密度估计。目前结合图像和点云完成异常检测任务基本上都是将二者特征进行拼接,导致训练模型的特征维度变得较高,高维特征特别是不同模态的特征之间会有一定的干扰性,导致异常检测的速度和正确率明显下降。

## 3 背景知识

### 3.1 无监督学习

无监督学习<sup>[9]</sup>是指不依赖于标签数据进行模型训练,而是试图挖掘数据的内在结构和关系。无监督学习主要包括聚类(将相似的数据点分为一组)、降维(将高维数据映射到低维空间)和异常检测(识别数据中的离群点)等任务。在大规模工业生产场景下,异常样本的数量相比于正常样本的数量要少很多很多,并且异常样本的异常原因也各不相同,无法穷举出异常的种类,故无监督学习更加适用于处理异常检测任务。

### 3.2 标准化流

Normalizing Flows(标准化流)的核心思想是对一个简单的已知分布(例如标准正态分布)进行一系列的可逆变换<sup>[10]</sup>,使其变为一个更复杂的分布,经过多次可逆变换最终与数据的实际分布相近,由于每一步的变换都是可逆的,所以对于每一步变换都存在一个逆变换,通过这些逆变换可以将数据的实际分布转换为标准正态分布。

## 4 方法设计

本文提出的 ADNF 方法以 RGB 图像和点云为输入,先将图像和点云的背景去除<sup>[11]</sup>,然后将点云的深度信息融合到 RGB 图像信息中,一般情况下的异常检测中,大部分的异常(例如颜色和形状等)都可以通过 RGB 图像进行识别,只有部分的凹陷和孔洞无法通过 RGB 图像进行识别,需要利用点云的空间深度信息进行辅助判断,故将图像和点云融合后可以利用二者的信息进行互补,从而进行更加准确的异常检测。然后使用预训练的 ResNet 特征提取网络并修改该网络的输入层,使该网络能够接受四通道的 RGBD 图像并提取到相应的特征,最后将提取到的特征用来训练标准化流模型并用于异常检测任务。如图 1 所示为本文提出的方法的总体框图,其中省略号表示重复以上过程。可以看到 ADNF 主要由 3 个重要部分组成:

(1)多模态融合:为了解决单一模态无法完整检测出各种异常的问题,本文提出了多模态融合,即将点云的空间深度信息与 RGB 图像进行融合,获取更加全面的信息,进而提高检测的准确性。(2)修改的预训练特征提取网络:针对图像的特征提取以及相关的特征含义的研究已经很成熟,故可以在之前的研究基础上挑选适合异常检测任务的预训练的特征提取网络并进行适当的修改,以提取出可以区分出正常和异常的特征用于训练标准化流模型。(3)标准化流模型:通过一系列的可逆变换将特征的分布转化为正态分布,将融合后的数据特征训练一个可逆的流模型用于异常检测任务,并根据模型输出的似然程度进行异常判别。

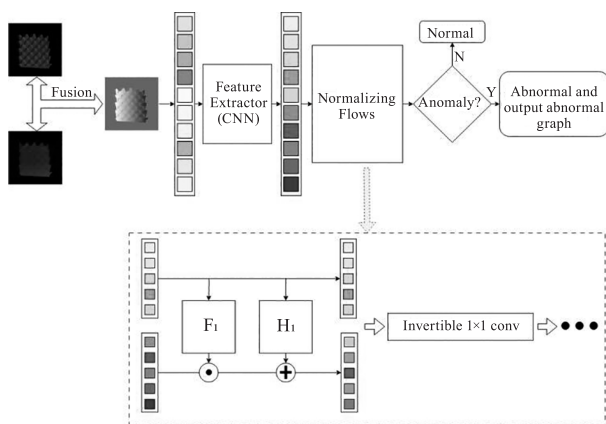


图 1 ADNF 完整流程示意图

Fig. 1 Schematic diagram of the complete process of ADNF

### 4.1 多模态融合

在深度学习的发展过程中,单一模态的处理方式已经被证明是有限的,特别是在异常检测的场景中。解决这一问题的有力方法是采用多模态融合的策略。多模态融合的核心思想在于将来自不同来源和性质的数据整合到一个统一的框架中,以使系统能够充分利用多维度的信息进行决策。如图 2 所示是将两种数据在通道中进行融合的过程。

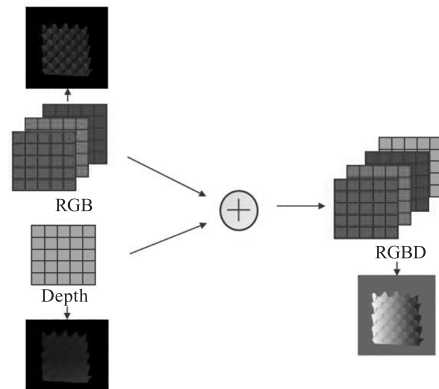


图 2 RGB 图像与点云融合过程

Fig. 2 RGB image and point cloud fusion process

### 4.2 预训练特征提取网络

在本文中,使用预训练的 wide\_resnet50\_2 网络模型作为基础,并根据输入的数据特点将模型的第一层卷积层替换为接收四个通道的输入,并将第四个通道权重设置为原前三个通道的平均值。ResNet<sup>[12]</sup>即残差神经网络(Residual Neural Network),其设计初衷是为了解决深层神经网络在训练时面临的梯度消失和梯度爆炸问题。其核心理念是引入残差块(Residual Block),通过短路连接(Shortcut Connection)来跳过一些层,实现前向和反向传播时更稳定和高效的信息传递。在特征提取时的作用主要体现在:通过深层网络捕捉复杂特征,利用深层结构可以更好地捕捉图像或其他数据中的高级特征。如图 3 所示是所使用的 wide\_resnet50-2 网络模型的残差结构。

### 4.3 标准化流模型

由修改后 wide\_resnet50\_2 网络模型所提取得到的高维特征将用来训练标准流模型,图 4 是 Normalizing Flows(标准化流)的理论示意图,Normalizing Flows(标准化流)的核心思想是对一个简单的已知分布进行一系列的可逆变换,使已知的简单分布逐步逼近数据的真实分布,也即是可以通过已知的简单分布表示未知的复杂分布。通过变量替换原理和雅克比

矩阵可以完整表示整个可逆变换过程。雅可比矩阵是一个描述多维变量可逆变换如何在局部变化时影响各维度的矩阵,具体的转换过程如下列公式所示:设  $z$  为一个多维随机变量, $\pi$  是  $z$  的概率密度函数, $F$  是一个多维函数, $x = F(z)$  是  $z$  的一个可逆变换, $\rho$  是  $x$  的概率密度函数。雅可比矩阵  $J_F$  定义为:

$$J_F = \begin{pmatrix} \frac{\partial x_1}{\partial z_1} & \dots & \frac{\partial x_1}{\partial z_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial x_n}{\partial z_1} & \dots & \frac{\partial x_n}{\partial z_n} \end{pmatrix} \quad (1)$$

转换的公式为:

$$p(x') \left| \det \begin{pmatrix} \frac{\partial x_1}{\partial z_1} & \dots & \frac{\partial x_1}{\partial z_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial x_n}{\partial z_1} & \dots & \frac{\partial x_n}{\partial z_n} \end{pmatrix} \right| = \pi(z') \quad (2)$$

代入得:

$$p(x') |\det(J_F)| = \pi(z') \quad (3)$$

两边同时除以行列式的值得:

$$p(x') = \pi(z') \left| \frac{1}{\det(J_F)} \right| \quad (4)$$

其中,  $\det$  表示的是计算行列式的值,  $z'$  表示任意取一个多维随机变量,  $x'$  则是  $z'$  通过  $x = F(z)$  变换得到的变量。行列式是一个标量值,提供关于线性变换影响体积的信息,当变量  $z$  通过  $x = F(z)$  进行变换得到变量  $x$  时,雅可比矩阵行列式的绝对值表明这个变换在局部如何拉伸或压缩空间。在标准化流中,空间的压缩或扩展会影响密度的大小<sup>[13]</sup>,使用雅可比矩阵行列式来调整这个密度,确保在变换后的空间中,概率密度函数的总和仍然为 1,从而确保整个变换是一个可逆的变换过程。

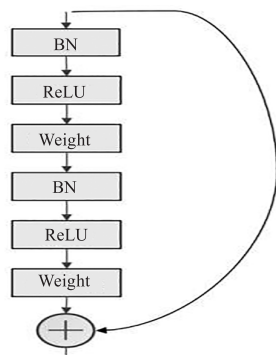


图3 wide\_resnet50\_2 网络模型的残差结构

Fig. 3 wide\_resnet50\_2 residual structure of network models

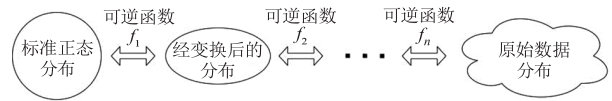


图4 Normalizing Flows 理论示意图

Fig. 4 Theoretical schematic diagram of Normalizing Flows

## 5 实验结果及比较

### 5.1 实验环境

为验证本文提出的 ADNF 算法有效性,设计了相关实验进行分析。实验硬件配置为 Intel(R) Xeon (R) Silver 4210R CPU @ 2.40 GHz 及 Quadro RTX 5000,实验软件配置为 Ubuntu 20.04.6 LTS、Python 3.8、Pytorch 1.9.0 及 CUDA Version 11.5。

### 5.2 实验数据

本文实验数据采用的数据集为 MVTEC 3D-AD 数据集的真实数据集,该数据集是用于无监督异常检测和定位的第一个全面的 3D 数据集,使用了高分辨率工业 3D 传感器获得不同对象类别的深度扫描来描述物体几何表面数据。数据集是由来自 10 个现实世界物体类别的 4147 个高分辨率 3D 点云扫描组成,训练集和验证集只包含无异常的数据,测试集中的样本包含正常的数据和各种类型的异常(按照对象样本在真实世界检查场景中可能出现的方式设计和制造的),且为每个异常提供精确的基础事实注释。

### 5.3 评价指标

#### 5.3.1 AUROC 指标

AUROC 指标,也称为接受者工作特征曲线下的面积(Area Under the Receiver Operating Characteristic Curve),是用于评估分类器性能的一种重要指标。Image AUROC 指标是评估模型在图像级别上区分正常图像和异常图像的能力。ROC 曲线是真正率(TPR)对假正率(FPR)的图示,真正率是正确识别为异常的异常图像的比例,假正率是错误地识别为异常的正常图像的比例。AUROC 是 ROC 曲线下的面积,可以量化地表达模型的性能。Image AUROC 值越高,表明模型在区分正常和异常图像方面的能力越好。Pixel AUROC 这个指标不是在整个图像级别上,而是在像素级别上评估模型。它用于衡量模型在识别图像中的异常像素(而不是整个图像)方面的效能。这对于那些需要精确定位图像中异常区域的任务尤其重要。像素级的 AUROC 计算方法与图像级别类似,但它是基于单个像素而不是

整个图像。像素级的 AUROC 值越高,表明模型在定位图像中的具体异常区域方面的能力越强。

### 5.3.2 AUPRO 指标

AUPRO 指标 (Area Under the Precision-Recall Curve) 即 PR 曲线下的面积。PR 曲线 (Precision-Recall Curve), 在不同阈值下, 以召回率 (Recall) 作为横坐标轴, 查准率 (Precision) 作为纵坐标轴绘制的曲线, 精确度是正确识别为异常的异常图像占所有识别为异常图像的比例, 而召回率是正确识别为异常的异常图像占所有实际异常图像的比例。对于不平衡的数据集 (即异常图像远少于正常图像), AUPRO 通常是一个更敏感的性能度量。AUPRO 的高值表明模型在维持较高精确度的同时实现了较高的召回率, 该指标是用于评估正样本不均衡情况下的模型性能。

### 5.4 结果对比

如下图 5 所示的异常分数图是经过标准化流模型得出的结果以及对应的 Ground Truth 图像, 可以观察到异常分数图显示出明显的异常区域, 并且通过与相应的 Ground Truth 图像进行对比可以发现, 异常图中的异常区域与真实异常位置基本一致, 这说明通过该方法可以有效地检测出异常。

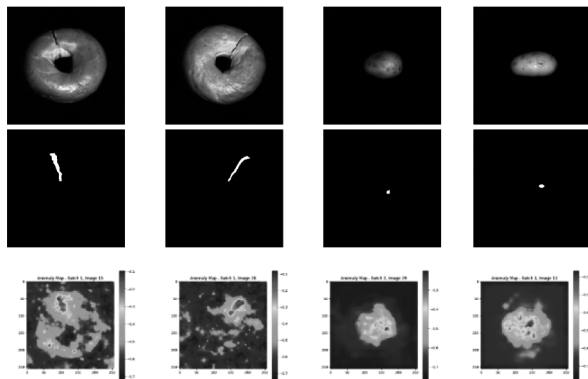


图 5 饼干和土豆的 RGB 图像、Ground Truth 图像及异常图

Fig. 5 RGB images, Ground Truth images, and anomaly images of cookies and potatoes

表 1 主要对比了不同的方法应用于 MVTEC 3D-AD 数据集中的 Image AUROC 指标, 其中 Depth GAN 通常指的是使用深度信息的生成对抗网络, 包括一个生成器和一个鉴别器, 生成器试图创建逼真的深度图像, 而鉴别器则试图区分真实图像和生成器产生的图像。Depth AE 是使用深度信息的自编码器, 通过一个编码器将数据压缩成低维表示, 然后

通过一个解码器重建数据。Depth VM 利用深度信息来模拟人类视觉系统对显著性区域的注意力分配。Voxel GAN 是一种应用于三维数据 (通常是体素化表示的数据) 的生成对抗网络。Voxel AE 是针对体素数据的自编码器, 可以有效地学习三维数据的压缩表示。Voxel VM 将视觉显著性模型应用于体素数据, 以识别三维空间中的显著性特征或区域。Depth iNet 是集成了深度信息的神经网络模型, 通过结合深度数据和传统的图像数据来提高识别、分类或检测任务的性能。Raw 通常指的是直接处理未经处理或最小处理的数据, 在图像处理中直接从相机传感器读取的数据。HoG 是一种特征描述符, 用于对象检测中, 通过计算图像局部区域的梯度方向直方图来描述这些区域的外观和形状。由于 MVTEC 3D-AD 数据集中的异常均为局部的小异常, 在加上模型训练数据集的数据量较小且具有一些噪声的影响, 导致了 Image AUROC 指标并没有很高, 但对比其他方法, 使用点云的深度信息辅助图像进行检测的 ADNF 方法的平均值为 77.8%, 相比于第二的 Depth iNet 提升了 8.4%。由于数据集有 10 种不同类别的对象, 并且各种类别的对象及其异常差异比较大, 本文所提出的 ADNF 方法表面规则的电缆接头和暗销等类别的检测中表现较优, 但是在表面有随机不规则纹路的曲奇饼干和马铃薯的异常检测中表现较差, 但相比于其他方法而言总体平均检测正确率有提升。

表 2 主要对比了不同方法的 AUPRO 指标, 用于评估正样本不均衡情况下的模型性能, 在 MVTEC 3D-AD 数据集中每一类的测试集中每一种异常类型的数据量为训练集的十分之一, 故可使用该指标评估模型的性能。从表 2 中的数值可以看出, ADNF 方法的平均值为 86.2%, 相比于第二的 HoG 提升了 9.1%。表 3 主要对比了不同方法的 Pixel AUROC 指标, 该指标通常表示异常分割的准确性, 从表 3 中可以看出, ADNF 方法的平均值为 95.8%, 相比于第二的 Depth iNet 提升了 2.6%。表 4 主要对比了仅使用 RGB、仅使用点云和使用融合后的数据的实验结果, 从三个指标的数值对比可以证明点云的空间深度信息对于异常检测的提升具有一定的作用。其中加粗的是最高的数值, 加下划线的是第二高的数值。

表1 各方法的 Image AUROC 指标对比

Tab.1 Comparison of Image AUROC metrics for various methods

| 方法                         | bagel        | cable gland  | carrot       | cookie       | dowel        | foam         | peach        | potato       | rope         | tire         | mean         |
|----------------------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
| Depth GAN <sup>[14]</sup>  | 0.538        | 0.372        | 0.580        | 0.603        | 0.430        | 0.534        | 0.642        | <u>0.601</u> | 0.443        | 0.577        | 0.532        |
| Depth AE <sup>[14]</sup>   | 0.648        | 0.502        | 0.650        | 0.488        | 0.805        | 0.522        | <b>0.712</b> | 0.529        | 0.540        | 0.552        | 0.595        |
| Depth VM <sup>[14]</sup>   | 0.513        | 0.551        | 0.477        | 0.581        | 0.617        | <u>0.716</u> | 0.450        | 0.421        | 0.598        | 0.623        | 0.555        |
| Voxel GAN <sup>[14]</sup>  | 0.680        | 0.324        | 0.565        | 0.399        | 0.497        | 0.458        | 0.566        | 0.579        | 0.601        | 0.482        | 0.517        |
| Voxel AE <sup>[14]</sup>   | 0.510        | 0.540        | 0.384        | 0.693        | 0.446        | 0.632        | 0.550        | 0.494        | 0.721        | 0.413        | 0.538        |
| Voxel VM <sup>[14]</sup>   | 0.553        | <u>0.772</u> | 0.484        | <u>0.701</u> | 0.751        | 0.578        | 0.480        | 0.466        | 0.689        | 0.611        | 0.609        |
| Depth iNet <sup>[15]</sup> | <u>0.690</u> | 0.597        | <b>0.753</b> | <b>0.862</b> | <u>0.881</u> | 0.590        | 0.597        | 0.598        | <u>0.791</u> | 0.577        | <u>0.694</u> |
| Raw <sup>[15]</sup>        | 0.627        | 0.507        | 0.600        | 0.654        | 0.573        | 0.524        | 0.532        | <b>0.612</b> | 0.412        | <u>0.678</u> | 0.572        |
| HoG <sup>[15]</sup>        | 0.487        | 0.587        | 0.691        | 0.545        | 0.643        | 0.596        | 0.516        | 0.584        | 0.507        | 0.430        | 0.559        |
| ADNF (Ours)                | <b>0.967</b> | <b>0.862</b> | <u>0.712</u> | 0.682        | <b>0.953</b> | <b>0.725</b> | <u>0.695</u> | 0.537        | <b>0.938</b> | <b>0.709</b> | <b>0.778</b> |

表2 各方法的 AUPRO 指标对比

Tab.2 Comparison of AUPRO indicators for various methods

| 方法                         | bagel        | cable gland  | carrot       | cookie       | dowel        | foam         | peach        | potato       | rope         | tire         | mean         |
|----------------------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
| Depth GAN <sup>[14]</sup>  | 0.421        | 0.422        | 0.778        | 0.696        | 0.494        | 0.252        | 0.285        | 0.362        | 0.402        | 0.631        | 0.474        |
| Depth AE <sup>[14]</sup>   | 0.432        | 0.158        | 0.808        | 0.491        | 0.841        | 0.406        | 0.262        | 0.216        | 0.716        | 0.478        | 0.481        |
| Depth VM <sup>[14]</sup>   | 0.388        | 0.321        | 0.194        | 0.570        | 0.408        | 0.282        | 0.244        | 0.349        | 0.268        | 0.331        | 0.335        |
| Voxel GAN <sup>[14]</sup>  | 0.664        | 0.620        | 0.766        | 0.740        | 0.783        | 0.332        | 0.582        | 0.790        | 0.633        | 0.483        | 0.639        |
| Voxel AE <sup>[14]</sup>   | 0.467        | 0.750        | 0.808        | 0.550        | 0.765        | 0.473        | 0.721        | <u>0.918</u> | 0.019        | 0.170        | 0.564        |
| Voxel VM <sup>[14]</sup>   | 0.510        | 0.331        | 0.413        | 0.715        | 0.680        | 0.279        | 0.300        | 0.507        | 0.611        | 0.366        | 0.471        |
| Depth iNet <sup>[15]</sup> | <u>0.763</u> | 0.676        | 0.884        | <b>0.883</b> | <u>0.864</u> | 0.322        | <b>0.881</b> | 0.840        | 0.844        | 0.634        | 0.759        |
| Raw <sup>[15]</sup>        | 0.402        | 0.314        | 0.639        | 0.498        | 0.251        | 0.259        | 0.527        | 0.531        | 0.808        | 0.215        | 0.444        |
| HoG <sup>[15]</sup>        | 0.712        | <u>0.761</u> | <u>0.932</u> | 0.487        | 0.833        | <u>0.520</u> | 0.743        | <b>0.949</b> | <u>0.916</u> | <b>0.858</b> | <u>0.771</u> |
| ADNF (Ours)                | <b>0.911</b> | <b>0.922</b> | <b>0.936</b> | <u>0.833</u> | <b>0.924</b> | <b>0.720</b> | <u>0.871</u> | 0.811        | <b>0.959</b> | <u>0.733</u> | <b>0.862</b> |

表3 各方法的 Pixel AUROC 指标对比

Tab.3 Comparison of Pixel AUROC indicators for various methods

| 方法                         | bagel        | cable gland  | carrot       | cookie       | dowel        | foam         | peach        | potato       | rope         | tire         | mean         |
|----------------------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
| Depth iNet <sup>[15]</sup> | <u>0.957</u> | 0.901        | 0.966        | <b>0.970</b> | <u>0.967</u> | 0.771        | <b>0.971</b> | <u>0.949</u> | 0.977        | 0.891        | <u>0.932</u> |
| Raw <sup>[15]</sup>        | 0.803        | 0.750        | 0.849        | 0.801        | 0.610        | 0.696        | 0.830        | 0.772        | 0.951        | 0.670        | 0.773        |
| HoG <sup>[15]</sup>        | 0.911        | <u>0.933</u> | <b>0.985</b> | 0.823        | 0.936        | <u>0.862</u> | 0.923        | <b>0.987</b> | <u>0.980</u> | <b>0.955</b> | 0.930        |
| ADNF (Ours)                | <b>0.976</b> | <b>0.980</b> | <u>0.982</u> | <u>0.957</u> | <b>0.980</b> | <b>0.879</b> | <u>0.963</u> | 0.947        | <b>0.992</b> | <u>0.925</u> | <b>0.958</b> |

表4 对比实验结果比较

Tab.4 Comparison of experimental results

| 方法    | Image AUROC  | AUPRO        | Pixel AUROC  |
|-------|--------------|--------------|--------------|
| 仅 RGB | 0.695        | 0.798        | 0.933        |
| 仅 3D  | 0.648        | 0.724        | 0.904        |
| ADNF  | <b>0.733</b> | <b>0.860</b> | <b>0.957</b> |

## 6 总结

本文提出了一种新颖的基于 RGB 图像和点云深度信息的异常检测方法,能够有效解决工业异常检测中数据利用不充分导致异常检测不到的问题。关键创新点在于引入了点云的空间深

度信息来辅助 RGB 图像进行异常检测并采用了标准化流模型来实现多源异构信息的有效融合。实验结果表明,相比于其他的方法,ADNF 方法可以显著提升异常检测的准确率。这充分证明了点云深度信息在提高异常检测性能方面的重要作用。本文的研究不仅提供了一种集成点云和 RGB 信息的有效方案,也展示了标准化流模型处理高维数据的强大能力。期待本研究能够对异常检测技术产生积极影响,并激发更多遵循这一思路的创新性研究。

## 参考文献:

- [1] Bergmann P, Jin X, Sattlegger D, et al. The MVTEC 3D-AD dataset for unsupervised 3D anomaly detection and localization[J/OL]. <https://arXiv.org/abs/2112.09045>.
- [2] Kang Zixiang, Wang Shengzhe, Cui Yuyong, et al. A voxelization laser point cloud object detection algorithm based on transformer[J]. *Laser & Infrared*, 2023, 53(2): 202 – 207. (in Chinese)  
康自祥, 王升哲, 崔雨勇, 等. 基于 Transformer 的体素化激光点云目标检测算法[J]. *激光与红外*, 2023, 53(2): 202 – 207.
- [3] Kingma D P, Dhariwal P. Glow: generative flow with invertible 1x1 convolutions [J/OL]. <https://arXiv.org/abs/1807.03039>.
- [4] Feng Xingming, Dai Yunfeng, Ding Yajie, et al. Anomaly detection method for underground cable pipelines based on reconstruction error[J]. *Automation and Instrumentation*, 2023, (9): 104 – 109. (in Chinese)  
冯兴明, 戴云峰, 丁亚杰, 等. 基于重构误差的地下电缆管道异常检测方法[J]. *自动化与仪器仪表*, 2023, (9): 104 – 109.
- [5] Wang G, Han S, Ding E, et al. Student-teacher feature pyramid matching for unsupervised anomaly detection [J]. *arXiv*, 2021; arXiv. 210304257.
- [6] Zhang Hong, Chen Zhaole, Dong Fangmin, et al. Image tampering detection based on feature fusion and local anomalies [J/OL]. *Computer Simulation* [2024 – 09 – 19]. (in Chinese)  
张虹, 陈赵乐, 董方敏, 等. 基于特征融合与局部异常的图像篡改检测[J/OL]. *计算机仿真*, 2024. [2024 – 09 – 19].
- [7] Zavrtnik V, Kristan M, SKOČAJ D. Reconstruction by inpainting for visual anomaly detection[J]. *Pattern Recognition*, 2021, 112: 107706.
- [8] Yu J, Zheng Y, Wang X, et al. FastFlow: unsupervised anomaly detection and localization via 2D normalizing flows [J/OL] <https://arXiv.org/abs/2111.07677v2>.
- [9] Cui Yulin, Wang Yanqiu, Liang Hongyu, et al. A method for extracting ToF point cloud targets based on unsupervised learning [J]. *Laser & Infrared*, 2022, 52(11): 1729 – 1736. (in Chinese)  
崔钰琳, 王艳秋, 梁红玉, 等. 基于无监督学习的 ToF 点云目标提取方法[J]. *激光与红外*, 2022, 52(11): 1729 – 1736.
- [10] Yoshibengio L D. Nice: Non-linear independent components estimation [J]. *arXiv e-prints*, 2014; arXiv 1410.8516.
- [11] Su Yunlong, Ping Xueliang, Li Nan. A plane extraction algorithm based on RANSAC 3D point cloud [J]. *Laser & Infrared*, 2019, 49(6): 780 – 784. (in Chinese)  
苏云龙, 平雪良, 李楠. 基于 RANSAC 三维点云的平面提取算法[J]. *激光与红外*, 2019, 49(6): 780 – 784.
- [12] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 770 – 778.
- [13] Dinh L, Sohl-Dickstein J, Bengio S. Density estimation using Real NVP [C]//International Conference on Learning Representations. 2016.
- [14] Bergmann P, Jin X, Sattlegger D, et al. The MVTEC 3D-AD dataset for unsupervised 3D anomaly detection and localization [C]//Proceedings of the 17th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications, 2022: 202 – 213.
- [15] Hrwitz E, Hoshen Y. Back to the feature: classical 3D features are (almost) all you need for 3D anomaly detection [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023: 2967 – 2976.