arXiv:2304.02216v2 [cs.CV] 1 Sep 2023

工业异常检测中的域偏移：一个真实世界数据集与掩码多尺度重建

张子龙[1] 赵志斌[1] 张兴武[1]*孙闯[1] 陈雪峰[1]西安交通大学机械工程学院[1]

zhangzilongc@gmail.com zhaozhibin@xjtu.edu.cn xwzhang@mail.xjtu.edu.cn
ch.sun@xjtu.edu.cn chenxf@xjtu.edu.cn

摘要

工业异常检测（*IAD*）对于自动化工业质量检测至关重要。数据集的多样性是开发全面工业异常检测算法的基础。现有工业异常检测数据集侧重于数据类别的多样性，而忽视了同一数据类别内领域的多样性。在本文中，为弥合这一差距，我们提出了航空发动机叶片异常检测（*AeBAD*）数据集，由两个子数据集组成：单叶片数据集和叶片视频异常检测数据集。与现有数据集相比，*AeBAD*具有以下两个特点：*1.)* 目标样本未对齐且处于不同尺度。*2.)* 测试集中正常样本的分布与训练集之间存在域偏移，其中域偏移主要由光照和视角的变化引起。基于此数据集，我们观察到当前最先进的（*SOTA*）工业异常检测方法在测试集中正常样本的领域发生偏移时表现出局限性。为解决此问题，我们提出了一种名为掩码多尺度重建（*MMR*）的新方法，该方法通过掩码重建任务增强模型推断正常样本中图像块之间因果关系的能力。*MMR*在*AeBAD*数据集上相比最先进的方法取得了更优的性能。此外，*MMR*在*MVTec AD*数据集上检测不同类型异常时，与最先进的方法取得了有竞争力的性能。代码和数据集可在以下地址获取 *https://github.com/zhangzilongc/MMR*。

1. 引言

工业异常检测（IAD）涉及识别和定位工业流程中的异常，范围从细微变化（如细划痕）到较大的结构性缺陷（如缺失组件） [4], ，且先验知识有限或完全没有关于异常的先验知识。工业异常检测有许多应用，例如确保生产高质量产品的智能制造流程，以及依赖机器人的自动化维护。由于工业流程通常经过优化，以尽量减少缺陷产品的产出，因此往往难以收集足够的异常样本用于训练 [42]。此外，由于工业流程会受到不可控随机因素的影响，可能产生不同类型的异常样本 [28, 49]。换言之，通常难以获得种类多样且数量充足的异常样本。因此，工业异常检测通常采用单类学习设定，仅使用正常 [1] 数据。

现有工业异常检测方法 [12, 21, 33, 35, 36, 44, 47] 主要依赖于 MVTec AD 等数据集 [4], VisA [52], 和 MVTec LOCO AD [3], ，这些数据集具有一些共同特征，例如样本经过对齐 [2] ，且每个类别内的尺度保持一致（如图1所示）。然而，在现实世界的自动化产品/组件检测场景中，这些条件未必总能成立。控制不准确或环境变化等因素可能导致物体尺度和视角出现显著差异 [19, 29]。这表明，现有数据集所提供的宽松条件可能无法充分反映现实世界工业异常检测应用的复杂性。因此，目前尚不清楚现有工业异常检测方法在如此复杂的条件下能否取得同等程度的成功，这也引发了关于工业异常检测技术未来发展与评估的重要问题。

为解决上述问题并弥合工业异常检测（IAD）实际应用与现有数据集之间的差距，我们在本文中提出了 **A**ero-**e**ngine **B**lade **A**nomaly**D**etection（AeBAD）数据集。AeBAD 的目标是自动检测叶片中的异常

*通讯作者

[1]本文中，“正常”按定义是“异常”的反义词。[2]VISA数据集中的一些其他物体图像描绘了未注册的物体。

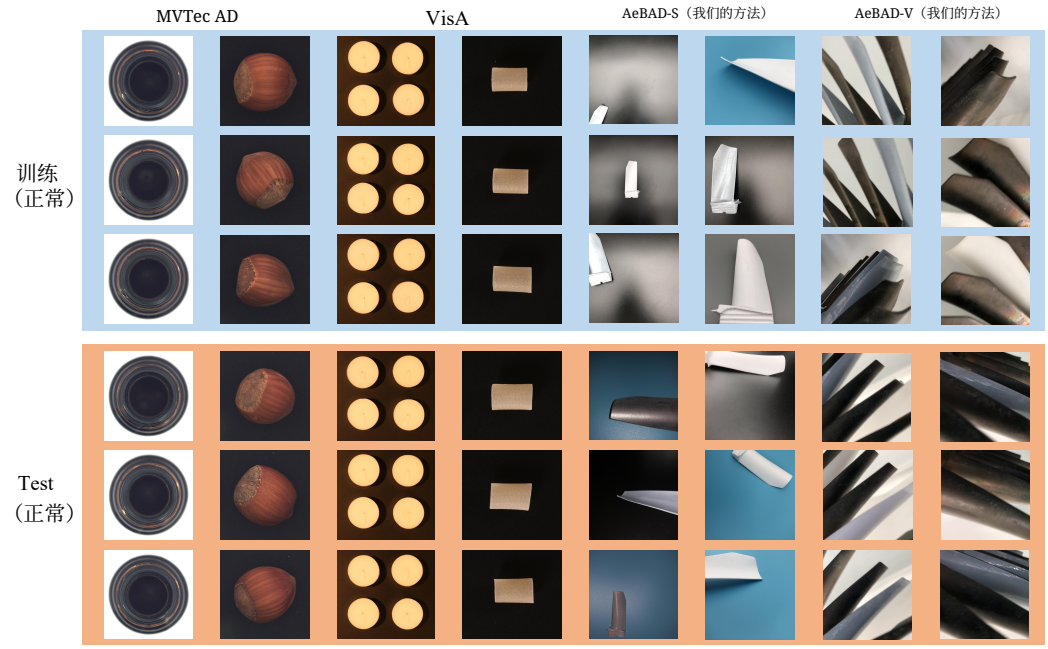

图1. 当前数据集和我们数据集中训练集与测试集的正常样本。所提出的AeBAD数据集的特点是：1.) 目标样本未对齐且处于不同尺度。2.) 测试集与训练集中正常样本的分布之间存在域偏移，其中领域偏移主要由光照和视角的变化引起。AeBAD数据集旨在自动检测航空发动机叶片的异常，以确保航空发动机的稳定运行。

航空发动机叶片，以确保其稳定运行。与以往侧重于检测缺陷类别多样性的数据集不同，AeBAD 聚焦于同一数据类别内领域的多样性。AeBAD 由两个子数据集组成：单叶片数据集（AeBAD-S）和叶片视频异常检测（AeBAD-V）。AeBAD-S 包含不同尺度的单叶片图像，其主要特征是样本未对齐。此外，测试集中正常样本的分布与训练集之间存在领域偏移，该领域偏移主要由光照和视角变化引起。部分示例如图1所示。另一方面，AeBAD-V 包含组装在航空发动机整体叶盘上的叶片视频，旨在检测叶盘旋转过程中的叶片异常。AeBAD-V 的一个显著特征是测试集中的拍摄视角与训练集不同。视频的部分帧如图1所示。AeBAD-V 与航空发动机叶片检测的实际需求高度一致，凸显了工业异常检测（IAD）的重要性。

构建 AeBAD数据集后，我们评估了最先进（SOTA）的工业异常检测方法，并得出以下观察结果。首先，依赖合成异常的方法在样本未对齐且尺度不同时表现不佳，因为大多数合成异常都分布在背景等非目标区域。这降低了区分正常样本与合成异常的难度，导致模型难以泛化到未见过的异常。其次，当测试集中的正常样本域发生偏移时，所有测试方法都会提高样本正常部分的异常分数，从而产生大量假阳性。这是因为一些现有方法严重依赖于对训练集中正常样本分布的建模；当分布发生变化时，训练集与测试集中正常样本的特征便会不匹配。我们认为，这一问题的根本原因是忽视了正常样本中的因果关系。

受上述观察和分析的启发，我们提出了一种名为**M**asked **M**ulti-scales**R**econstruction (MMR) 的方法，用于从掩码输入重建完整图像的多尺度特征。我们的方法无需合成异常，从而规避了上述问题之一。同时，MMR的前置任务增强了模型对正常样本中因果关系的感知，从而提高了对领域偏移的鲁棒性。与现有的图像修复方法相比，我们的方法侧重于重建不同尺度的特征而非繁杂的细节，从而使输出具有更高的语义清晰度。此外，我们的掩码策略与以往不同，其中掩码部分被完全隐藏。我们将证明这种掩码策略可以

显著提高异常检测的性能。我们的实验表明，MMR在减少假阳性和提升AeBAD数据集上的异常检测性能方面优于最先进的（SOTA）方法。此外，MMR在MVTec AD数据集上实现了与SOTA方法相当的性能。

本文的主要贡献如下：

1. 我们认识到工业异常检测的实际应用与现有数据集之间存在差距，因此提出了一个名为 AeBAD 的新数据集。与以往的数据集不同，AeBAD 考虑了不同尺度和视角等多种实际因素，这会导致测试集中正常样本发生域偏移。该数据集与航空发动机叶片异常检测的需求高度契合，凸显了工业异常检测的实际意义。据我们所知，这是首个考虑领域偏移的工业异常检测数据集，因此更能代表真实世界的场景。

2. 我们发现，当测试集中正常样本的领域发生偏移时，现有的工业异常检测最先进的方法会显现出局限性。为解决这一问题，我们提出了一种名为 MMR 的新方法，该方法增强了模型推断正常样本中图像块之间因果关系的能力。在 AeBAD 数据集上，我们的方法取得了优于最先进方法的性能。此外，MMR 在 MVTec AD 数据集上的性能也能与最先进方法相媲美。

本文其余部分组织如下。第2节回顾相关工作。第3节介绍我们的数据集，并指出当前方法的不足。第4节介绍所提出的5 节展示实验结果。第6 节对与 MMR 密切相关的因素进行消融分析。第7 节介绍我们工作的局限性及未来工作。第8节给出结论。

2. 相关工作

我们首先回顾现有的工业异常检测数据集，然后概述相关的工业异常检测方法。更全面的工业异常检测综述见 [9, 20, 39]。本文不回顾与语义异常相关的方法和数据集。

2.1. 数据集

当前深度学习方法的成功在很大程度上依赖于数据分布的多样性。这种多样性涵盖大量数据类别，例如 ImageNet [34], ，以及同一数据类别的不同域，例如 NICO [50] 和 OOD-CV [51]。在不同域的数据集上评估目标算法的性能，是将其应用于真实场景的重要步骤。尽管目前已有多个包含各类工业产品、用于工业异常检测任务的数据集，例如 MVTec AD [4], 、DAGM [40], 、KSDD [38] 和 Sensum-SODF [31], ，但据我们所知，这些数据集都忽略了训练集与测试集之间存在的域偏移，而这种情况在裂纹分割 [25], 、利用机器人视觉进行铁路轨道维护 [32], 、表面缺陷检测 [37]等任务中均存在。解决这一问题并考虑域偏移，对于开发更稳健、更有效且能可靠应用于实际场景的工业异常检测算法至关重要。

Bergmann 等人 [4]提出了 MVTec 异常检测数据集（MVTec AD），该数据集侧重于检测细微变化和较大的结构性缺陷。MVTec AD 的主要特点是所有物体大致对齐，如图1所示。Zou 等人 [52] 提出了视觉异常（VisA）数据集，该数据集涵盖更复杂的结构，并在单张图像中呈现多个物体。最近提出的 MVTec 逻辑约束异常检测数据集（MVTec LOCO AD） [3] 旨在检测逻辑异常及结构性异常。此外，Bergmann 等人 [6] 还推出了一个用于工业异常检测的 3D 数据集。

然而，现有数据集都忽略了训练集与测试集之间存在的域偏移。受航空发动机叶片异常检测实际应用的启发，本文提出了一种新型数据集，其中包含训练集与测试集之间的域偏移。这些域偏移涵盖不同的视角和光照条件，所提出的数据集能够弥补当前数据集的部分不足。

2.2. 方法

尽管工业异常检测中存在多种设置，例如将正常数据与少量噪声数据（异常）混合使用， [41],本文着眼于实际应用，仅评述工业异常检测中的单类学习方法，即训练时仅有正常数据可用。根据异常检测的形式化方式，我们将这些方法分为三类。

第一类将异常检测表述为匹配问题。这些方法基于正常数据建立模式，并在测试期间将测试数据的模式与已建立的模式进行匹配。例如，PaDiM[10] 建立正常数据每个图像块的分布，并计算测试数据的似然性以检测异常。类似地，FastFlow [45] 和 CFLOW-AD [14]对正常数据的归一化流进行建模。PatchCore[33] 构建多尺度特征的“数据库”，并通过k近邻搜索匹配测试数据。MKDAD

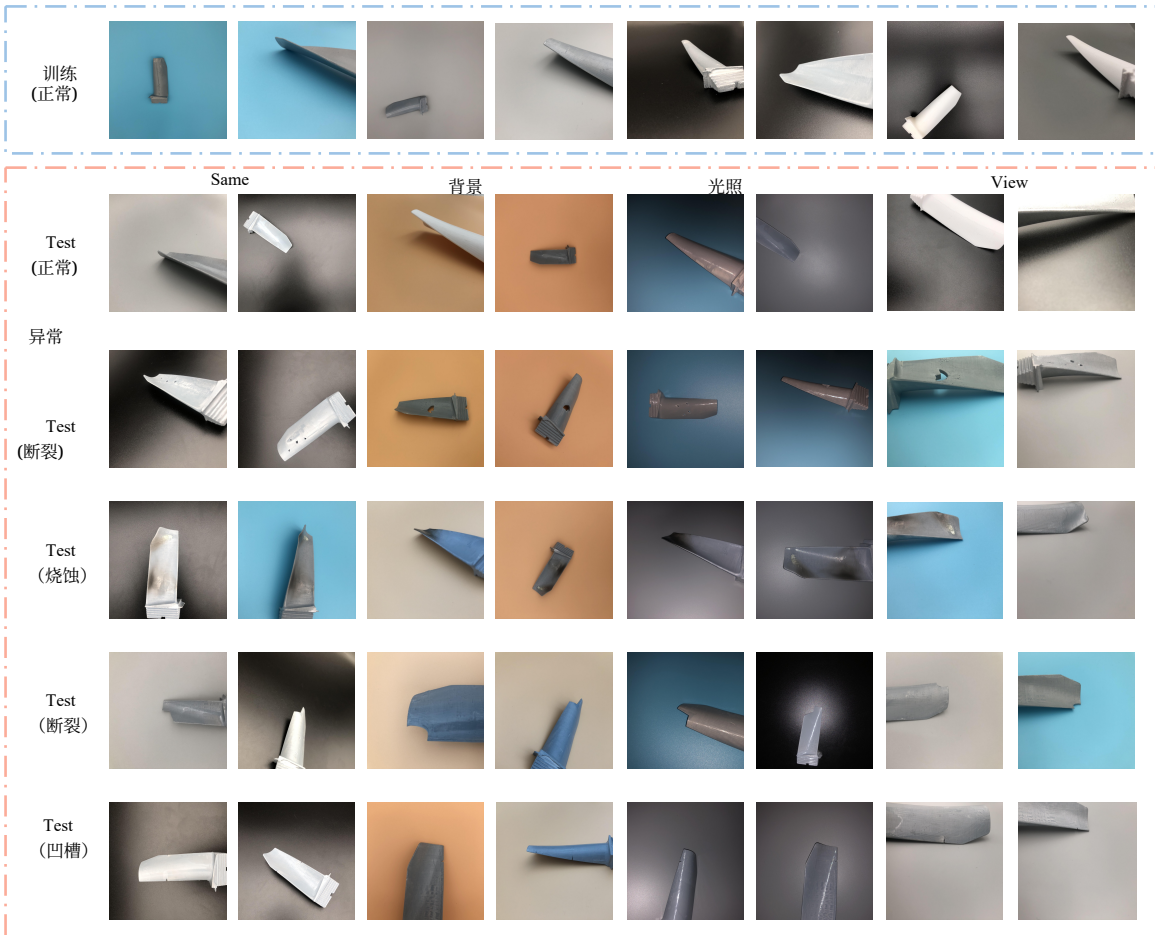

图 2. AeBAD-S 数据集上的训练集和测试集样本。训练集仅由正常样本组成。测试集包含四个类别（每两幅图为一类），分别为相同、背景、光照和视角。在“相同”类别中，测试集正常样本的分布与训练集相同。“背景”类别涉及图像背景的变化，而“光照”类别涉及光照条件的变化。“视角”类别涉及拍摄位置的变化，新位置与训练集中的位置对称。此外，叶片样本中的异常分为四类：断裂、烧蚀、断裂和凹槽。凹槽类别指叶片边缘的微小缺陷，而断裂类别指叶片缺失的区域。（**最佳彩色视图**）

[35]和反向KD [12]将教师网络生成的一般模式与学生网络生成的正常数据模式进行匹配以检测异常。

第二类方法希望模型能够将区分正常样本与合成异常的能力泛化到区分未见异常上。这些方法的主要区别在于合成异常的构造方式。例如，CutPaste [21]将某个物体的部分区域裁剪下来，再贴到其他物体上以合成异常。NSA [36]使用泊松图像编辑，将一个图像中的物体无缝克隆到另一张图像中，从而显著减少不连续现象。DRAEM[47]则根据佩林噪声生成的掩码和一个额外的数据集来生成异常。

第三类包括基于重建的方法。这些方法 [7, 11, 27] 依赖这样一种假设：仅在正常样本上训练的重建模型，只能在正常区域中成功，而在异常区域中则会失败 [44]。这种思路的主要问题在于，模型可能学会一些技巧，从而也能很好地修复异常。为解决这一问题，研究者采用了不同的策略。RIAD [48] 和 InTra [30] 提出了一种掩码图像修复前置任务。由于图像的某些部分被掩码，模型需要感知图像中的因果关系，才能恢复被掩码部分的信息，从而避免采用平凡解。UniAD [44] 提出了邻域掩码编码器和逐层查询解码器，以防止“恒等捷径”。

与我们工作最相关的方法是 RIAD [48]和 InTra [30]。主要区别如下。

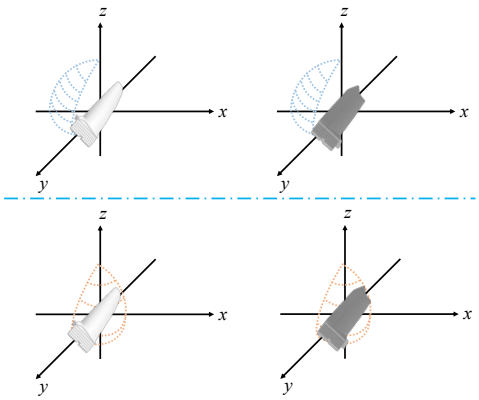

图3。**顶部**：AeBAD-S 上训练集中的采集图像的拍摄位置。所有拍摄位置均从蓝色点线的位置中采样。**底部**：“视角”测试集中采集图像的拍摄位置。所有拍摄位置（橙色点线）均从与蓝色点线位置对称的位置中采样。（**最佳彩色视图**）

首先，我们不对整幅图像中的复杂细节进行重建，而是重建不同尺度的特征。这使模型的输出包含更多语义信息。其次，我们采用的掩码策略有所不同：掩码部分会被完全隐藏，因此其中的特征也会被彻底移除；而在以往的方法中，这些特征会与可见部分的特征共享。

3. 航空发动机叶片异常检测数据集

3.1. 背景与意义

航空发动机是飞机的心脏，其叶片健康状况对于确保稳定运行至关重要。航空发动机叶片的缺陷可能会严重影响发动机效率，甚至导致其故障 [22]。由于成本高昂且发生频率低，很难获得各种具有不同缺陷（异常）的叶片。因此，叶片缺陷检测通常被视为单类异常检测问题。在真实世界的叶片检测场景中，孔探仪器采集的叶片图像可能因孔探仪器在航空发动机上的位置不同而在视角和光照方面有所差异。由于待检测的航空发动机未知，其视角也未知。这促使我们探索测试集域偏移情况下的单类工业异常检测。为满足真实应用需求，实验中的叶片要么来自真实发动机，要么是 1:1 复制品。

表1. AeBAD数据集的统计概览。AeBAD-V中的视频已被分割成连续帧。“N”表示正常样本的数量，“A”表示异常样本的数量。

	训练	Test							
AeBAD-S	521	same		背景		光照		view	
		N	A	N	A	N	A	N	A
		230	459	93	212	75	198	92	280
AeBAD-V	707	视频-1		视频-2			视频-3		
		N	A	N	A			N	A
		533	368	251	650			381	520

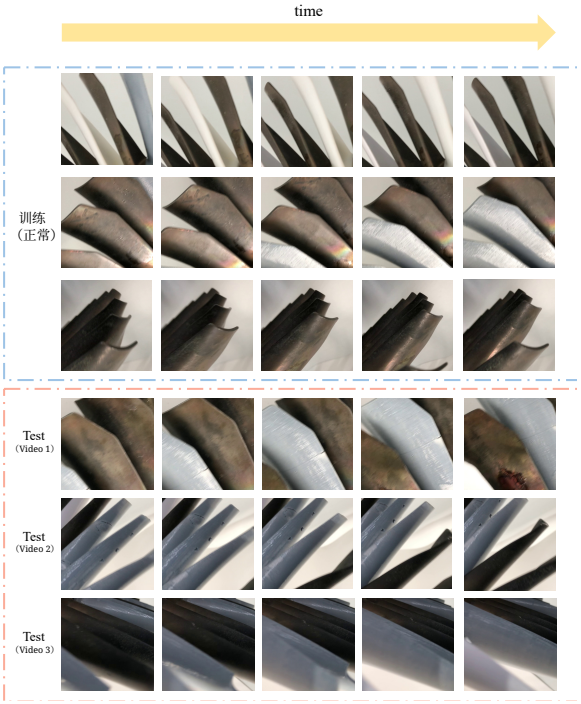

图4. AeBAD-V 数据集中训练集和测试集视频的帧。从左到右，每一行展示了随时间变化的一系列视频帧。与 AeBAD-S 不同，一张图像中有多个叶片。测试集的拍摄视角与训练集不同。

3.2. 数据集描述

为了更好地满足真实场景中叶片异常检测的需求，并弥合当前数据集与实际应用之间的差距，我们推出了**航**空**发**动**机**叶片**异**常**检**测（AeBAD）数据集，该数据集由两个子数据集组成：AeBAD-S 和 AeBAD-V。AeBAD-S 包含单个叶片的图像，而 AeBAD-V 则包含航空发动机整体叶盘上多个叶片的视频。AeBAD 的训练集仅包含正常样本，而测试集既包含正常样本，也包含异常样本。叶片异常包括以下四种缺陷：断裂、烧蚀、凹槽和断裂。断裂由异物（例如，

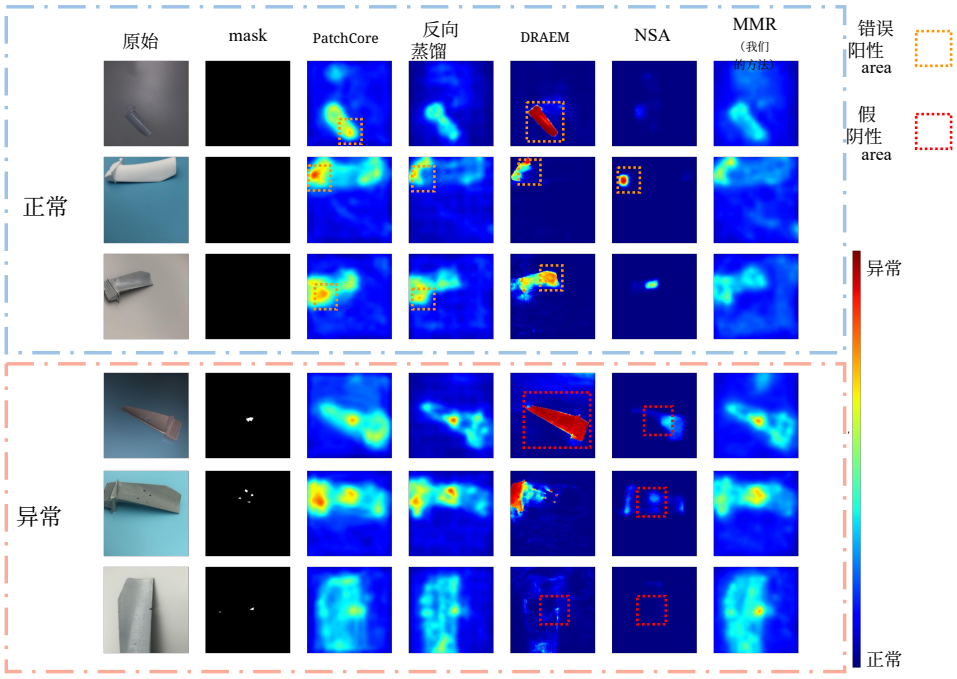

图 5. 当前最先进的方法和 MMR 的异常图（热力图）可视化。红色区域被视为异常区域，其他颜色视为正常区域。随着领域偏移，PatchCore 和 RecerseDistillation 错误地将正常区域评估为异常区域。由于图像的异常分数是该图像生成的热力图的最大值，这使得 PatchCore 和 RecerseDistillation 产生一些假阳性。相反，DRAEM 和 NSA 漏检较小的缺陷，从而产生一些假阴性区域。对于我们提出的 MMR，它可以在检测异常的同时减少假阳性。（**建议彩色查看**）

沙子、金属、鸟类、冰雹等）通常会导致叶片断裂。[43]烧蚀是由叶片所处的高温气体造成的。烧蚀会降低叶片的性能。当烧蚀达到一定程度，且叶片材料的性能无法满足要求时，就会发生断裂。凹槽是由大量弹性应力循环造成的，并会不断扩展，最终导致叶片断裂。部分异常样本如图2所示。AeBAD 的统计概况列于表1中。训练样本总数为 1228 个，测试样本总数为 4342 个。此外，对于 AeBAD-S 数据集中出现的所有异常，我们均提供像素级真实标注；对于 AeBAD-V 数据集中的所有视频帧，我们均提供样本级标注。

与当前所有图像大致对齐的数据集不同，AeBAD-S 训练集中采集图像的拍摄位置取自蓝色虚线所标示的位置，这些位置由 x 轴负半轴、y 轴正半轴和 z 轴正半轴围成。拍摄位置如图 3上方所示。此外，训练图像有三种背景颜色：黑色、灰色和蓝色。部分训练样本如图 2第一行所示。AeBAD-S 测试集分为四类：相同、背景、光照和视角。在“相同”类别中，测试集正常样本的分布与训练集中的分布相同。此外，“背景”类别涉及图像背景的变化，“光照”类别涉及光照条件的变化。最后，“视角”类别涉及拍摄位置的变化，新位置与训练集中的位置对称。“视角”测试集的拍摄位置如图 3 下方所示。在上述测试集中，除“相同”测试集外，其余测试集均存在域偏移。不同测试集的样本如图 2所示。

AeBAD-V 包含从不同视角拍摄的视频，训练集由从四个视角拍摄的视频组成，而测试集则包含从与训练集不同的三个视角拍摄的视频。部分视频帧如图 4 所示。与处于模拟环境中的 AeBAD-S 相比，AeBAD-V 中的环境更符合航空发动机叶片的真实异常检测。

3.3. 当前方法的不足

我们提出的 AeBAD 数据集使我们能够评估当前工业异常检测方法在领域偏移下的性能。我们在 AeBAD-S 数据集上评估了四种最先进的方法：PatchCore [33], ReverseDistillation[12], DRAEM [47], 以及 NSA [36]。需要指出的是

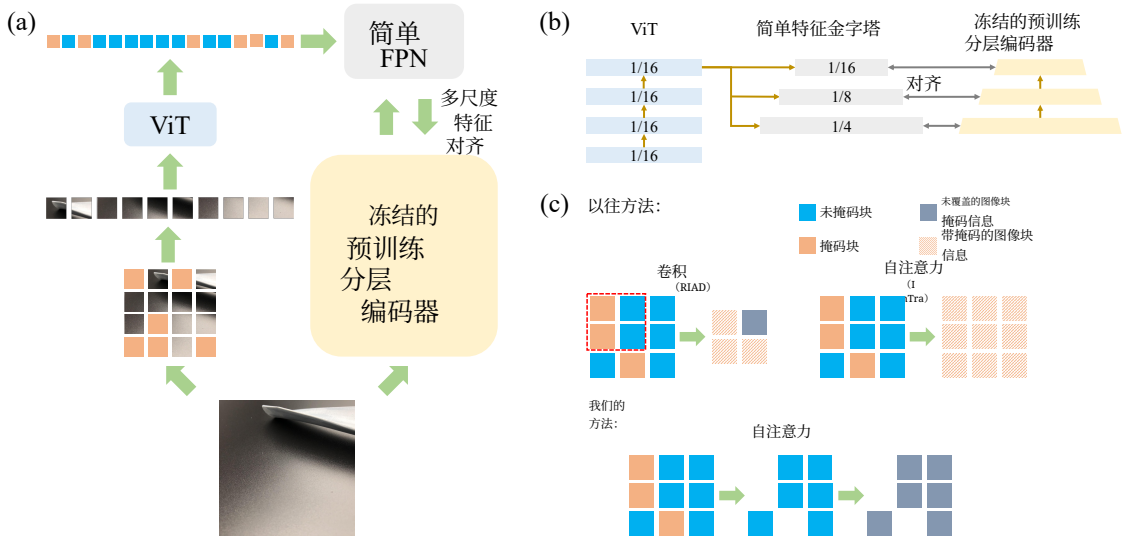

图6. **(a)** 掩码多尺度重建（MMR）的整体框架。MMR由两部分组成。 **(b)** Simple FPN示意图 [23]. “ $_{1\ 16}$”表示视觉Transformer中每个图像块的感受野为 16×16。 **(c)** 以往方法与我们方法的区别。红色虚线表示卷积核。以往方法通过卷积或自注意力，将掩码区域的特征与可见区域的特征共享。我们的方法可防止可见区域的信息泄漏。（**最好以彩色查看**）

在这些方法中，图像的异常分数是该图像生成的热力图中的最大值。

上述方法的定性结果如图5所示，更多结果见图9。在定性结果中，红色区域被视为异常区域，其他颜色则被视为正常区域。我们发现以下问题：1.) 基于合成异常的方法（例如NSA）往往会漏检缺陷，并产生假阴性区域。对于DRAEM，其失效原因更为复杂，我们将在第5节进行分析。2.) 基于匹配的方法，如PatchCore和ReverseDistillation，在领域偏移下往往会将正常区域误判为异常区域，从而产生许多假阳性。

关于问题1，原因在于样本未对齐且尺度不同，因此合成异常大多分布在背景区域。这显著降低了区分正常样本与合成异常的任务难度，导致模型对未见异常的泛化能力较差。关于问题2，现有方法严重依赖于对训练集中正常样本分布的建模，这会在域发生偏移时导致训练集与测试集中正常样本的特征存在差异。我们认为，问题2很可能是由于未考虑正常样本中的因果关系所致。

4. 掩码多尺度重建

受上述观察与分析的启发，我们提出一种无需合成异常的异常检测方法，称为**掩**码式**多**尺度**重**建（MMR）。MMR 的整体框架如图6 （a）所示。

4.1. MMR 详情

MMR由两部分组成。第一部分如图6 (a)左侧所示，将图像均匀划分为若干图像块（图6中为16块），并对其中一定比例的图像块进行掩码处理，其余部分保留；所有图像块的高和宽均相等。将保留的图像块输入视觉Transformer（ViT） [13], , 以生成这些图像块的嵌入表示。随后，将随机初始化的掩码块嵌入追加到保留图像块的嵌入之后。使用简单特征金字塔网络（Simple FPN） [23]接收所有图像块的嵌入作为输入，并生成多尺度特征。第二部分如图6(a)右侧所示，使用冻结的预训练分层编码器，将第一部分使用的相同输入编码为多尺度特征。MMR的目标是对齐这两部分生成的多尺度特征。恢复掩码图像块信息的根本动机是，希望模型能够感知图像不同部分空间位置之间的依赖关系。在MMR中，重建任务负责关注细节，而掩码恢复任务则用于建模各部分之间的因果关系。

Simple FPN 的架构 [23] Simple FPN 遵循 [1, 24]中的思路。其目标是构建一个包含多尺度特征的特征金字塔。与以往需要将不同尺度的特征作为输入并需要侧向连接的工作 [24], 不同,

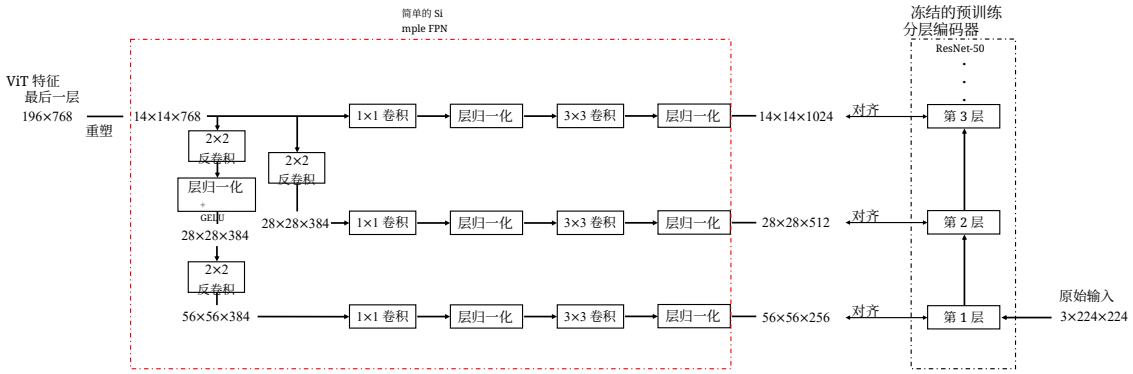

图 7. 简单的 Simple FPN 架构 [23]（**如红色虚线框所示**）。假设原始输入的维度为 $3 \times 224 \times 224$。ViT 中的图像块大小为 16×16，视觉Transformer 的特征维度为 768。图中展示了特征金字塔包含 $\times 4$、$\times 2$ 和 $\times 1$ 特征的情况。假设冻结的预训练分层编码器为 ResNet-50。

简单 FPN 只需使用视觉Transformer 最后一层的特征，即可并行解码出不同尺度的特征。示意图如图6 (b) 所示。我们采用 Simple FPN 的原始架构 [23], , 具体如下：与视觉Transformer 最后一层特征尺度相同的尺度使用视觉Transformer 的最终特征图。尺度 $\times 2$ (或 $\times 4$) 由一层（或两层） 2×2 步幅为=2的反卷积层构建。在$\times 4$ 尺度情况下，第一层反卷积后接 LayerNorm（LN） [2] 和 GELU [17]。然后，对于每个金字塔层级，Simple FPN 会先应用一个带 LN 的 1×1 卷积，再应用一个同样带 LN 的 3×3 卷积。特征图的维度与冻结的预训练分层编码器中不同尺度特征的维度相同。Simple FPN 的架构如图7所示。

问题表述 形式化地说，在第一部分中，为了处理二维图像，我们首先将原始图像 $\mathbf{x} \in h \times w \times c$ 重塑为一系列展平的二维图像块 $\mathbf{x}_p \in n \times (p^2 \times c)$，其中 (h, w) 是原始图像的分辨率， c是通道数， $n = \frac{hw}{p^2}$ 是图像块数， (p, p) 是图像块的分辨率。然后，我们使用操作 $\Phi(; \eta)$ 随机提取$n_{masked} = \lfloor n \times (1 - \eta) \rfloor$ 个展平图像块作为未掩码块，其中 η 是掩码比例。之后，我们使用视觉Transformer $f_{ViT}(\cdot)$ 对未掩码块进行编码，其中视觉Transformer的每一层都由一个多头自注意力模块和一个前馈网络组成。随机初始化的掩码块嵌入通过操作 $\Psi(\cdot)$附加到视觉Transformer的输出 $f_{ViT}(\Phi(\mathbf{x}_p;\ \eta)) \in n_{masked} \times (p^2 \times c)$之后。最后，将所有图像块的嵌入$\Psi(f_{ViT}(\Phi(\mathbf{x}_p;\ \eta))) \in n \times (p^2 \times c)$重塑为 $\sqrt{n} \times \sqrt{n} \times (p^2 \times c)$，此时特征图上每个特征的感受野为 (p, p)。简单特征金字塔网络 $f_{FPN_i}(\cdot)$ 将所有嵌入解码为不同尺度的特征 $\mathbf{z}_{masked_i}$，其中 i 表示第 i 个尺度。在第二部分中，冻结的预训练层次编码器 $f_{FPH}(\cdot)$ 将$\mathbf{x}$编码为不同尺度的特征$\mathbf{z}_{frozen_i} \in (h_i \times w_i) \times c_i$，其中 h_i, w_i, c_i 分别表示高度、宽度和通道数。并且$\mathbf{z}_{masked_i}$ 的维度与$\mathbf{z}_{frozen_i}$相同。注意， h_i 和 w_i 是 $2^{i-1} \times \sqrt{n}$。MMR的最终损失函数如下：

$$\mathcal{L} = \sum_{i=1}^{s} \sum_{k=1}^{h_i \times w_i} \frac{1}{h_i \times w_i} \left(1 - \frac{\mathbf{z}_{masked_i}(k)(\mathbf{z}_{frozen_i}(k))^T}{||\mathbf{z}_{masked_i}(k)||||\mathbf{z}_{frozen_i}(k)||}\right), \tag{1}$$

其中 s 表示总尺度数，$\mathbf{z}_{masked_i}(k) \in 1 \times c_i$表示 $\mathbf{z}_{masked_i}$ 的第 k 行。我们的目标是最小化式（1），以恢复掩码块的多尺度特征。

在测试阶段，MMR 将原始**未掩码**输入（整幅图像）送入上述两个部分。第一部分和第二部分分别生成特征$\mathbf{z}_{masked_i}$ 和$\mathbf{z}_{frozen_i}$。每个尺度 i 的异常图 $\mathrm{AM_i} \in \mathrm{h_i} \times \mathrm{w_i}$ 计算如下：

$$\mathrm{AM_i} = \zeta((\mathbf{z}'_{\mathrm{masked_i}} \odot \mathbf{z}'_{\mathrm{frozen_i}})\mathbb{1}), \tag{2}$$

其中，$\mathbf{z}'_{masked_i}$ 是按行归一化后的$\mathbf{z}_{masked_i}$ 的结果，$\odot$ 表示哈达玛积， $\mathbb{1} \in c_i \times 1$ 是一个所有元素均为 1 的向量， $\zeta(\cdot)$ 是将向量 $v \in (h_i \times w_i) \times 1$ 重塑为 $v' \in h_i \times w_i$ 的操作。对于分辨率不同的特征图，均通过双线性插值统一上采样至原始图像的分辨率。**异常图：** 最终异常图 AM 是各上采样后 $\mathrm{AM_i}$ 的总和。**异常分数：** 一幅图像的异常分数为 AM 上的最大值。

MMR 的机制 在训练阶段，MMR 使用正常样本作为输入，并根据未掩码块提供的信息恢复掩码块的多尺度特征。这使视觉Transformer 和简单 FPN 能够

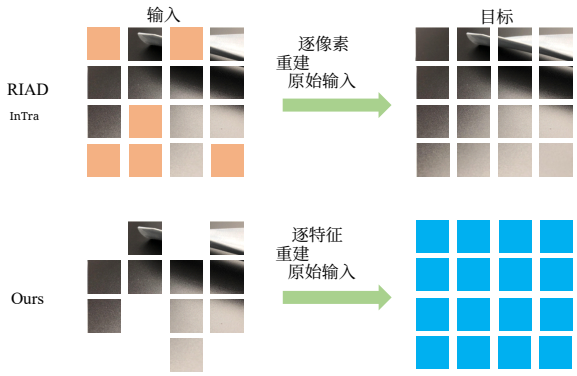

图 8. 以往方法与我们的方法重建目标的差异。

捕捉正常样本不同部分之间的因果关系，并避免学习对异常检测无益的恒等函数。这里，不同部分之间的因果关系指各部分空间位置之间的依赖关系。具体而言，人只需捕捉叶片根部的一部分，就能预测叶片的整体位置及其上的一些细节。在 MMR 中，这一过程被替换为根据未掩码块预测掩码块。从数学上讲，这一过程等价于建模条件分布，其中未掩码部分是条件，掩码部分是目标。如第3.3节所述，PatchCore 和反向蒸馏旨在对正常样本的分布（无条件分布）进行建模。当发生域偏移时，这会导致正常样本在训练集和测试集中的特征出现差异。与这两种方法不同，MMR 对不同部分之间的条件分布进行建模。发生域偏移时，尽管图像的分布发生变化，图像块之间的条件分布仍相对不变。因此，MMR 对域偏移相对不敏感。

测试期间，该方法的两个部分均使用原始非掩码输入。由于冻结的预训练分层编码器包含丰富的语义信息，因此能够感知未见过的异常。而对于 ViT 和 Simple FPN，由于它们只对正常样本中的因果关系进行建模，因此无法很好地重建异常部分。这会导致异常的重建多尺度特征和编码多尺度特征产生偏差，这些偏差对应于特征图中的异常部分。

4.2. MMR 与当前方法的关系

与图像修复方法的关系 与当前基于重建的方法相比，MMR主要有两点不同。首先，如图8所示，MMR并不试图重建完整图像的所有细节，而是专注于恢复不同尺度的特征。由于冻结的编码器对某些变化不敏感 [8],编码特征更加鲁棒。这使得MMR能够强调正常样本不同部分之间的因果关系。其次，与RIAD和InTra等先前方法相比，MMR使用了不同的掩码策略，如图6 (c)。在MMR中，掩码部分被完全隐藏，而在RIAD [48]（图 6 (c)中的卷积）和InTra [30]（图 6 (c)中的视觉Transformer），掩码部分的特征与可见部分的特征共享。MMR的方法防止可见部分的信息泄漏到掩码部分。这可以通过第6节中的烧蚀研究进一步验证。

与学生-教师框架的关系学生-教师框架，也称为知识蒸馏[18]，是一种处理单类异常检测任务的高效方法。当前方法，例如 MKDAD[35] 和 ReverseKD [12], 对正常样本的预训练特征进行了蒸馏。MMR 也可以被视为一个学生-教师网络，其中冻结的预训练编码器充当教师，另一部分则是学生网络。MMR 与当前知识蒸馏方法之间的独特区别在于，知识蒸馏中包含一个因果推理模块。这使模型能够感知图像中不同部分的空间位置，从而提升其在域偏移场景下的泛化能力。

与掩码自编码器（MAE）的关系 [15] MAE 采用了与 MMR 相同的掩码策略。尽管掩码策略相同，但其目的截然不同。MAE 使用这种掩码策略，即仅在少量图像块上训练，以降低内存消耗，从而允许使用更大的批次大小；而 MMR 使用这一策略来帮助模型感知图像中不同部分空间位置之间的依赖关系，并对正常样本中的因果关系进行建模。更重要的是，我们对这种掩码策略提出了新的见解：它可以防止可见部分的信息泄漏到掩码部分（如图6 （c）所示）。MMR 与 MAE 的另一处区别在于重建目标：MAE 重建原始的逐像素输入（类似于图8的顶部），而 MMR 执行逐特征重建。

4.3. 实现细节

输入图像经过随机裁剪增强后，调整大小至 256×256，然后进行中心裁剪至 224×224。图像块的分辨率为 (16, 16)。掩码比例 η 为 0.4。ViT 的主干网络采用原始 ViT-B [13], ，其主干网络由掩码自编码器进行预训练 [15]。Simple FPN 生成感受野分别为 4、8 和 16 的特征图。我们使用广泛采用的 WideResNet50

表 2. AeBAD-S 数据集上的样本级异常检测性能（AUROC %）。粗体表示最佳结果。报告了 5 次试验的结果。

方法	Same	背景	光照	View	Mean
PatchCore [33] CVPR' 2022	75.2 ± 0.3	74.1 ± 0.3	74.6 ± 0.4	60.1 ± 0.4	71.0
ReverseDistillation [12] CVPR' 2022年	82.4 ± 0.6	84.3 ± 0.9	85.5 ± 0.9	71.9 ± 0.8	81.0
DRAEM [47] 国际计算机视觉大会' 2021年	64.0 ± 0.4	62.1 ± 6.1	61.6 ± 2.7	62.3 ± 0.9	62.5
NSA [36] 欧洲计算机视觉会议' 2022年	66.5 ± 1.4	48.8 ± 3.5	55.5 ± 3.2	55.9 ± 1.1	56.7
RIAD [48] PR' 2020年	38.6 ± 0.6	41.6 ± 1.3	46.8 ± 0.8	33.0 ± 0.6	40.0
InTra [30] ICIAP' 2022年	39.8 ± 0.8	46.1 ± 0.5	44.7 ± 0.3	46.3 ± 1.5	44.2
MMR (本文)	**85.6** ± 0.5	**84.4** ± 0.7	**88.8** ± 0.5	**79.9** ± 0.6	**84.7**

表 3. AeBAD-S 数据集上的像素级异常检测性能（区域重叠率（PRO）%）。粗体表示最佳结果。报告了 5 次试验的结果。

方法	相同（区域重叠率（PRO））	背景（区域重叠率（PRO））	光照（区域重叠率（PRO））	视角（区域重叠率（PRO））	平均值（区域重叠率（PRO））
PatchCore [33] CVPR' 2022	89.5 ± 0.2	89.4 ± 0.1	88.2 ± 0.1	84.0 ± 0.2	87.8
ReverseDistillation [12] CVPR' 2022	86.4 ± 0.4	86.4 ± 0.7	86.7 ± 0.5	82.9 ± 0.7	85.6
DRAEM [47] 国际计算机视觉大会' 2021年	71.4 ± 4.2	44.3 ± 11.6	67.6 ± 2.7	71.1 ± 2.3	63.6
NSA [36] 欧洲计算机视觉会议' 2022年	43.0 ± 1.3	29.7 ± 2.1	59.9 ± 1.3	51.1 ± 0.1	45.9
RIAD [48] PR' 2020年	71.9 ± 1.3	33.4 ± 0.6	65.3 ± 1.0	62.2 ± 1.7	58.2
InTra [30] ICIAP' 2022年	76.8 ± 0.2	74.8 ± 0.3	73.7 ± 0.3	73.4 ± 0.2	74.7
MMR（我们的方法）	**89.6** ± 0.2	**90.1** ± 0.2	**90.2** ± 0.2	**86.3** ± 0.3	**89.1**

[16, 46] 作为冻结的预训练分层编码器，然后提取第 1、2 和 3 层的多尺度特征。我们使用 AdamW 优化器 [26] ($\beta_1 = 0.9$, $\beta_2 = 0.95$)，并采用逐步学习率衰减。学习率为 0.001。对于所有数据集，我们将训练轮数设为 200，批次大小设为 16。此外，所有实验均在 Ubuntu 18.04.5 上进行，使用的计算机配备 Xeon(R) Gold 6140R CPU@2.30GHZ 和一块具有 11GB 显存的 NVIDIA GeForce RTX 2080 Ti。

5. 实验

5.1. 数据集与对比方法

我们在包含 AeBAD-S 和 AeBAD-V 的 AeBAD数据集上评估 MMR。此外，我们还在 MVTec 数据集上评估 MMR，以展示其检测不同类型异常的能力。对比方法包括最新的 SOTA 方法以及与我们工作最相关的方法，包括 PatchCore [33], ReverseDistillation[12], DRAEM [47], NSA [36], RIAD [48] 和 InTra [30]。对于 PatchCore、ReverseDistillation、DRAEM 和 NSA，我们使用官方代码，其中 PatchCore 和 ReverseDistillation 的骨干网络为预训练的 WideResNet50。我们调整超参数以确保获得最佳性能。对于 RIAD 和 InTra，由于它们未提供官方代码，我们使用这些方法的非官方复现版本 [3]。

5.2. 结果

AeBAD-S： 指标：为评估样本级异常检测，我们采用受试者工作特征曲线下面积（AUROC），这是一种广泛使用的指标，它同时考虑假阳性和假阴性。对于像素级异常检测，我们采用区域重叠率（PRO） [5] 作为评估指标。PRO 分数对任意大小的异常一视同仁。与常见的像素级 AUROC 不同，如果未能检测到彼此不连通的异常区域，PRO 分数会显著下降。

定量结果见表2和3。定性结果见图9。我们可以观察到以下现象：1.) 除 DRAEM 外，所有方法在“视角”子集上的性能都显著低于其他子集。这主要是因为视角变化会增加正常样本的假阳性，如图5 和9所示。在图9异常样本部分的第三行中，MMR 也为正常区域生成了较高的分数。2.) 基于预训练模型的方法（如 PatchCore、ReverseDistillation 和 MMR）对背景和光照变化具有较强的鲁棒性。此外，与 Patc hCore 和 ReverseDistillation 相比，随着领域偏移加剧，我们提出的方法能够降低正常区域的异常分数，如图9正常部分所示。请注意，图像的异常分数是该图像生成的热力图中的最大值。PatchCore 和 ReverseDistillation 在“视角”数据集上产生了更多假阳性区域，这也是它们在该数据集上的 AUROC 性能较差的原因，尽管它们的 PRO 性能与 MMR 接近。3.) 与其他方法相比，NSA 在异常分类和定位方面性能较差，且方差较大。这主要是因为所提出的数据集未对齐，而且增强后的合成异常会显著影响性能。当合成异常较少分布在背景区域时，性能较好；反之亦然。

[3] RIAD 和 InTra 的非官方重新实现版本分别位于此处和此处。

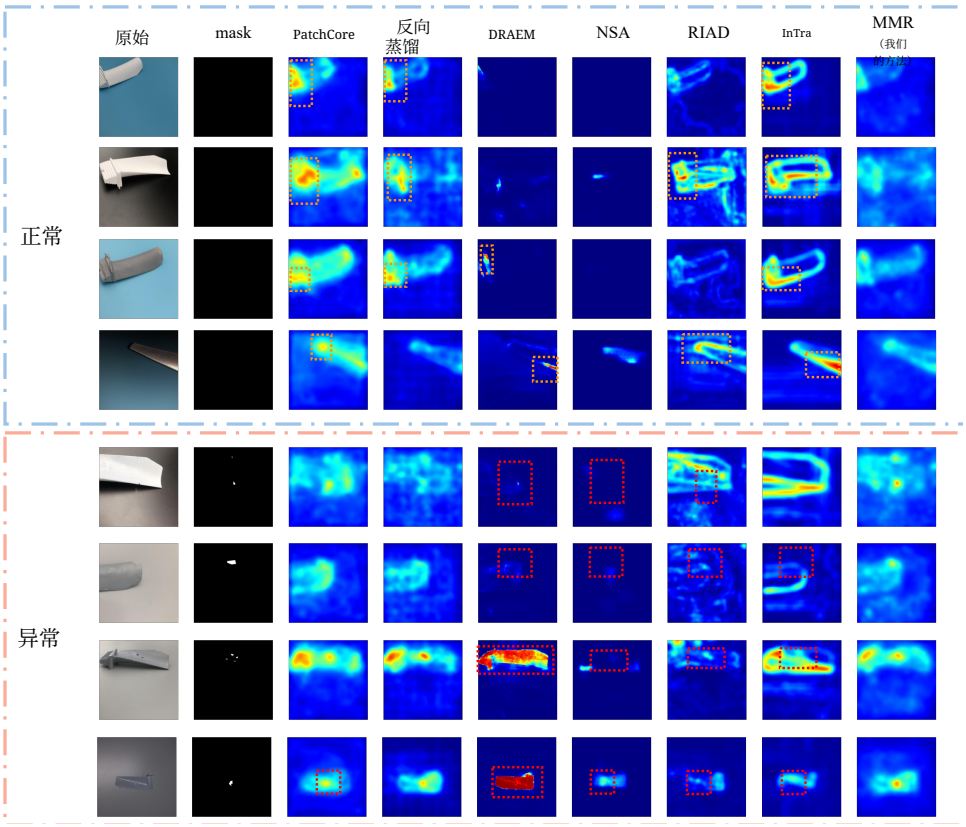

图 9. AeBAD-S 数据集上当前 SOTA 方法、最相关工作和 MMR 的异常图（热力图）可视化。虚线框和热力图的含义与图 5 中相同（**彩色视图效果最佳**）

。4.) 对于 DRAEM，由于它会在整幅图像中生成合成异常，因此能够检测出一些明显的异常（例如，图5异常部分的第二行）。但是，当光线和背景发生变化时，它会将整片叶片视为异常（例如，图5和9异常样本部分的第一行和第四行）。我们认为，主要原因是 DREAM 的两个大型模型导致其对数据过拟合，因此对存在域偏移的数据泛化能力较差。5.) RIAD 和 InTra 使用像素级重建，因此复杂细节会产生较大的重建误差。

此外，尽管在测试阶段，MMR 第一部分的输入是整幅图像，但该部分的输出并不与冻结的预训练分层编码器的输出相同。从图 5 和图9可以看出，MMR 能够检测出一些微小缺陷（异常）。这一结果表明，MMR 的第一部分能够避免“恒等捷径”。

AeBAD-V：指标：我们仅使用 AUROC 评估样本级异常检测。

定量结果见表 4。定性结果见图 10。与之前的结果类似，我们观察到视角发生变化时，部分

表 4. AeBAD-V 数据集上的样本级异常检测性能（AUROC %）。粗体表示最佳结果。报告了 5 次试验的结果。

方法	Video 1	Video 2	Video 3	Mean
PatchCore [33]	71.1 ± 0.2	86.0 ± 0.4	55.1 ± 0.7	70.7
反向蒸馏 [12]	66.0 ± 2.1	84.8 ± 1.8	62.1 ± 1.1	71.0
DRAEM [47]	**79.5** ± 5.0	71.2 ± 3.3	53.6 ± 1.2	68.1
NSA [36]	59.4 ± 3.5	72.7 ± 1.3	61.9 ± 4.0	64.6
RIAD [48]	78.0 ± 1.2	47.1 ± 2.9	43.2 ± 1.1	56.1
InTra [30]	62.7 ± 1.4	55.8 ± 1.3	43.7 ± 1.4	54.1
MMR（我们的方法）	75.7 ± 0.2	**88.3** ± 1.4	**70.7** ± 0.6	**78.2**

尤其是在 Video 2 和 Video 3 中，PatchCore 和反向蒸馏产生了更多假阳性，而 DRAEM 和 NSA 产生了更多假阴性。我们提出的方法 MMR 能够降低正常区域的异常分数，并检测出多个异常区域。值得注意的是，与 AeBAD-S 相比，AeBAD-V 中的物体更多分布在中心区域。这使 NSA 在某些情况下能够定位明显的异常，但 DRAEM 仍无法预测异常。我们认为，这主要也是 DREAM 过拟合所致。对于 DRAEM 在 Video 1 上的结果，由于 Video 1 与训练集具有相似的视角，DRAEM 能够取得良好的结果。

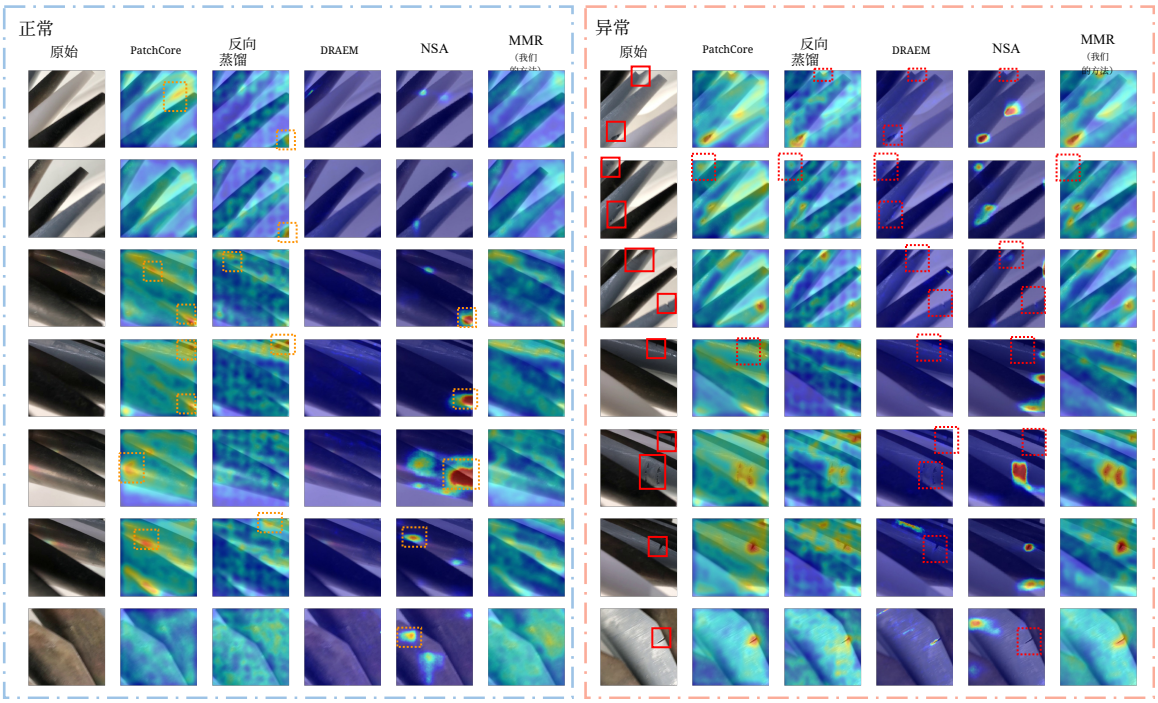

图10. 当前最先进方法、最相关研究及 AeBAD-V 数据集上 MMR 的异常图（热力图）可视化。虚线框和热力图的含义与图5中的相同。红色实线框表示异常区域的边界框。（**最佳彩色视图**）

此外，除了上述结果，我们还展示了连续视频帧（视频）的定性结果，详见 https://github.com/zhangzilongc/MMR。

MVTec： 评估指标：我们遵循通用协议。对于样本级异常检测，我们使用 AUROC。对于像素级异常检测，我们使用区域重叠率（PRO）和像素级 AUROC。此外，我们遵循常见设置，在每个产品类别上进行训练和测试。

定量结果见表5和6。我们可以观察到，MMR 同样取得了与 SOTA 方法具有竞争力的结果。这表明 MMR 能够检测不同类型的异常。值得注意的是，与 Ae-BAD 数据集上的优异性能相比，MMR 在 MVTec 数据集上的性能不够理想。主要原因是，当测试样本的分布发生变化时，MMR 能更好地区分正常样本和异常样本，而其他方法会受到域偏移的影响。然而，对于 MVTec 数据集，测试样本的分布与训练集一致，因此 MMR 的优势无法凸显。此外，我们观察到，与 PatchCore 相比，MMR 获得的异常掩码较为粗糙，这也是其像素级性能不够理想的原因。

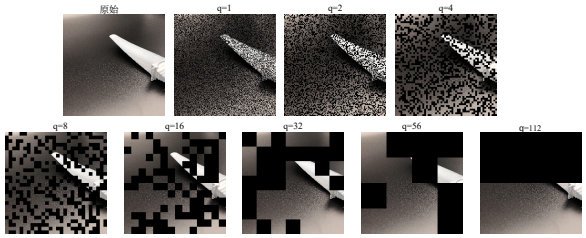

图11. 掩码比例相同的不同输入形式。q 指掩码的最小单位（$q \times q$ 像素）。

6. 烧蚀研究

在本部分中，我们研究了一些对 MMR 影响较大的因素。我们选择 AeBAD-S 数据集进行烧蚀研究。如果没有额外说明，实验的设置和超参数与第 4.3 节中的相同。

6.1. 输入形式

在 MMR 中，我们将未掩码补丁作为输入。在本部分中，我们选择整幅图像（所有图像块）作为输入，同时保持掩码比例不变。我们以 $q \times q$ 个像素为单位，然后以此为基准随机掩盖相当于整幅图像 $\eta = 0.4$ 倍的面积。我们选择 $q = [1, 2, 4, 8, 16, 32, 56, 112]$。不同掩码图

	PatchCore [33]	RD [12]	DRAEM [47]	NSA [36]	RIAD [48]	InTra [30]	MMR（我们的方法）
瓶子	100	100	99.2	97.7	99.9	100	100
电缆	99.5	95.0	91.8	94.5	81.9	84.2	97.8
胶囊	98.1	96.3	98.5	95.2	88.4	86.5	96.9
地毯	98.7	98.9	97.0	95.6	84.2	98.8	99.6
Grid	98.2	100	99.9	99.9	99.6	100	100
榛子	100	99.9	100	94.7	83.3	95.7	100
皮革	100	100	100	99.9	100	100	100
金属螺母	100	100	98.7	98.7	88.5	96.9	99.9
Pill	96.6	96.6	98.9	99.2	83.8	90.2	98.2
螺钉	98.1	97.0	93.9	90.2	84.5	95.7	92.5
Tile	98.7	99.3	99.6	100	98.7	98.2	98.7
牙刷	100	99.5	100	100	100	99.7	100
晶体管	100	96.7	93.1	95.1	90.9	95.8	95.1
Wood	99.2	99.2	99.1	97.5	93.0	98.0	99.1
Zipper	99.4	98.5	100	99.8	98.1	99.4	97.6
Mean	**99.1**	98.5	98.0	97.2	91.7	95.9	98.4

表 5. MVTec 上的样本级异常检测性能（AUROC %）。

	PatchCore [33]	RD [12]	DRAEM [47]	NSA [36]	RIAD [48]	InTra [30]	MMR（我们的方法）
瓶子	98.6/96.1	98.7/96.6	99.1/-	98.3/-	98.4/-	97.1/-	98.3/96.0
电缆	98.5/92.6	97.4/91.0	94.7/-	96.0/-	84.2/-	93.2/-	95.4/87.2
胶囊	98.9/95.5	98.7/95.8	94.3/-	97.6/-	92.8/-	97.7/-	98.0/94.5
地毯	99.1/96.6	98.9/97.0	95.5/-	95.5/-	96.3/-	99.2/-	98.8/96.6
Grid	98.7/95.9	99.3/97.6	99.7/-	99.2/-	98.8/-	99.4/-	99.0/96.5
榛子	98.7/93.9	98.9/95.5	99.7/-	97.6/-	96.1/-	98.3/-	98.5/91.2
皮革	99.3/98.9	99.4/99.1	98.6/-	99.5/-	99.4/-	99.5/-	99.2/98.6
金属螺母	98.4/91.3	97.3/92.3	99.5/-	98.4/-	92.5/-	93.3/-	95.9/88.6
Pill	97.6/94.1	98.2/96.4	97.6/-	98.5/-	95.7/-	98.3/-	98.4/96.1
螺钉	99.4/97.9	99.6/98.2	97.6/-	96.5/-	89.1/-	99.5/-	99.5/97.6
Tile	95.9/87.4	95.6/90.6	99.2/-	99.3/-	89.1/-	94.4/-	95.6/90.2
牙刷	98.7/91.4	99.1/94.5	98.1/-	94.9/-	98.9/-	99.0/-	98.4/93.0
晶体管	96.4/83.5	92.5/78.0	90.9/-	88.0/-	87.7/-	96.1/-	90.2/79.1
Wood	95.1/89.6	95.3/90.9	96.4/-	90.7/-	85.8/-	90.5/-	94.8/88.9
Zipper	98.9/97.1	98.2/95.4	98.8/-	94.2/-	97.8/-	99.2/-	98.0/95.0
Mean	**98.1**/93.5	97.8/**93.9**	97.3/-	96.3/-	94.2/-	97.0/-	97.2/92.6

表 6. MVTec 上像素级异常检测性能（像素级 AUROC % / 区域重叠率（PRO））。

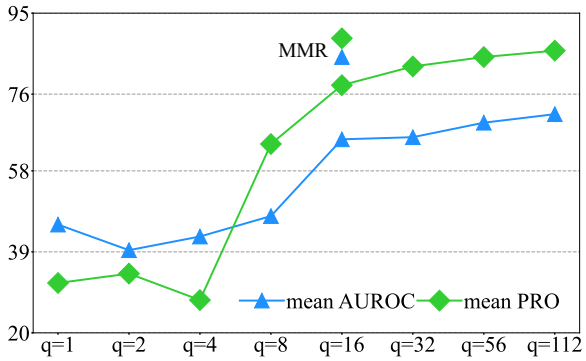

图12. AeBAD-S 数据集上不同最小单元的性能（平均 AUROC（%）和平均区域重叠率（PRO）（%））。

像示例 q 如图 11所示。请注意，当$q = 16$时，非掩码输入所包含的信息与 MMR 的原始输入相同。唯一的区别是，未掩码部分的信息是共享的，如图 6 （c）所示。不同输入的结果如图 12所示。我们可以观察到，当掩码比例保持不变时，不同掩码形式的输入会对性能产生显著影响。当 $q \leqslant 8$时，性能较差。原因虽然掩码补丁的信息未知，但由于单元较小，相邻区域会泄漏信息。这会使模型失去从全局上下文推断掩码补丁信息的能力。当 $q = 16$ 时，由于单元更大，性能有所提升。然而，由于这仍无法阻止信息泄漏，$q = 16$ 处的性能低于所提出的 MMR 在 $q = 16$处的性能。此外，我们发现了一个有趣的现象：随着最小单元不断增大，性能会持续提升。我们认为，主要原因是最小单元变大后，图像块之间的关联会越来越弱，即使发生信息泄漏，也不会影响整体重建。然而，部分重建结果的细节会变差，导致其平均区域重叠率（PRO）大幅上升，而平均 AUROC 基本没有提升。这与图13中的 0.9 掩码比例相似。

6.2. 掩码比例

我们通过采用不同的掩码比例进行实验，以探究其如何影响模型感知因果关系的能力。图13 展示了掩码比例从 0 到 0.9 时的分析结果。

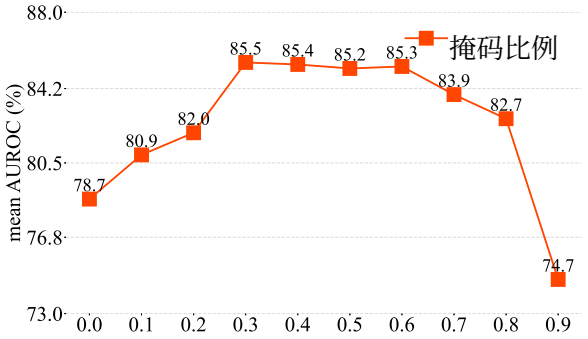

图13. 不同掩码比例在 AeBAD-S 数据集上的平均 AUROC（%）。

表7. 不同类型的预训练性能。PIMAE 表示视觉Transformer骨干网络通过掩码自编码器在 ImageNet 上进行预训练。

	平均 AUROC（%）	平均值区域重叠率（PRO）（%）
从头开始训练	75.1	88.5
PIMAE	84.7	89.1

我们的发现表明，模型的性能在掩码比例的两端都较低，而在中间区间达到最高。当掩码比例较小时，模型能够根据包含大量相关信息的周围区域，准确推断被掩码的图像块。然而，随着掩码比例增大，模型仅凭少数可见图像块推断完整图像的能力会减弱，因此性能得分也会降低。需要注意的是，这与近期的掩码自编码器 [15], 不同，后者的最佳掩码比例较高，目标是感知语义；工业异常检测模型则需要识别一些微小缺陷，因此需要较小的最佳掩码比例，才能准确重建细节。

6.3. 预训练类型

在我们的实现中，视觉Transformer骨干网络通过掩码自编码器在 ImageNet 上进行预训练 [15]。本节展示了从头开始训练视觉Transformer的结果，见表 7。直观来看，使用正常样本从头开始训练视觉 Transformer，应该能使模型更善于识别未见过的异常。然而，我们发现事实并非如此。主要原因是，在小型数据集上从头开始训练视觉Transformer可能导致过拟合，进而降低模型的泛化能力。此外，选择由掩码自编码器预训练的视觉Transformer还有一个优势：其预训练的上游任务与我们提出的下游任务大致相同，两者都旨在重建原始输入。这种相似性使模型能够更有效地迁移所学表征，从而在我们的任务上取得更好的性能。

表 8. 不同预训练分层编码器的性能。该预训练编码器在 ImageNet 1K 上进行预训练。

	平均 AUROC (%)	平均值区域重叠率（PRO）（%）
随机初始化的 ResNet-18	51.8	52.8
预训练的 ResNet-18	82.4	90.7
WideResNet-50	84.7	89.1

表 9. 不同尺度特征对性能的影响（平均 AUROC (%) 和平均区域重叠率（PRO）(%)）。

	第1层	第2层	第3层	平均值 AUROC (%) / 平均值区域重叠率（PRO）(%)
1	√			66.5 / 85.6
2		√		76.9 / 86.6
3			√	76.5 / 82.4
4	√	√		74.7 / 86.9
5		√	√	**85.0** /88.1
6	√	√	√	84.7 / **89.1**

表10. 当前最先进的方法和 MMR 的吞吐量（图像/秒）。

	PatchCore [33]	RD [12]	MMR（我们的方法）
吞吐量（图像/秒）	76.0	**190.4**	118.1

6.4. 冻结的预训练分层编码器

本节中，我们选择不同的冻结的预训练层次化编码器来训练 MMR。结果如表 8所示。我们可以观察到，MMR 对 ResNet 18 也具有良好的泛化能力。

6.5. 多尺度特征

在本部分中，我们报告不同尺度的特征对性能的影响。结果如表 9所示。我们可以观察到，中间尺度的特征更有利于检测缺陷，而其他尺度的特征则可以作为补充，提升性能。

7. 局限性与未来工作

MMR 的主要局限在于，推理测试图像时需要两个基础模型（一个冻结的预训练编码器和一个经过特定训练的模型），这会占用大量内存，从而限制其在实际应用中的使用。此外，所提出的基于因果变换器的模块对内存的需求较高，也妨碍了其实际应用。我们计算了部分 SOTA 模型和 MMR 的吞吐量，实验条件见第 4.3 节，输入图像为 224×224。结果列于表 10。我们可以观察到，由于 MMR 包含两个基础模型，其吞吐量低于 ReverseDistillation。未来的工作中，我们将探索基于卷积模块的因果模块，并将其与预训练模型集成到单个模型中。

此外，所提出的数据集仅关注单一物体（叶片），无法充分展现

该模型的异常检测能力。我们将探索更多实际的异常检测场景，其中将涵盖更多缺陷类型。

8. 结论

本文旨在介绍 AeBAD，这是一个用于解决域偏移问题的新数据集，而这一问题在现有数据集中尚未得到充分考虑。通过对该数据集进行分析，我们发现，当测试集中正常样本的域发生偏移时，当前最先进的方法存在不足。我们还探究了这些不足背后的原因，并提出了一种名为 MMR 的新方法。该方法增强了模型推断正常样本中图像块之间因果关系的能力，并提升了模型在不同领域偏移下的性能。此外，MMR 在检测各种类型的异常方面也具有竞争力。我们的工作为异常检测领域提出了一个新方向，使其更加贴近真实的工业场景。

致谢

本研究得到国家自然科学基金项目（项目编号：52175114&92060302）、燃气轮机科学中心项目（P2022-DC-I-003- 001）以及陕西省高层次人才特殊支持计划的资助。

参考文献

[1] Edward H Adelson, Charles H Anderson, James R Bergen, Peter J Burt, and Joan M Ogden. 图像处理中的金字塔方法。 *RCA*工程师, 29(6):33–41, 1984. 7[2] Jimmy Lei Ba, Jamie Ryan Kiros, and Geoffrey E Hin- ton. 层归一化。 *arXivpreprint arXiv:160706450*, 20168.. [3] Paul Bergmann, Kilian Batzner, Michael Fauser, David Sattlegger, and Carsten Steger. 超越凹痕和划痕：无监督异常检测和定位中的逻辑约束。 国际计算机视觉杂志, 130(4):947–969, 2022年. 1,3[4] Paul Bergmann, Michael Fauser, David Sattlegger, and Carsten Steger. MVTec AD——一种用于无监督异常检测的全面真实世界数据集。在*IEEE/CVF*计算机视觉与模式识别会议论文集, 第9592–9600页, 2019. 1,3[5] Paul Bergmann, Michael Fauser, David Sattlegger, and Carsten Steger. 不知情的学生：具有判别性潜在嵌入的师生异常检测。在*IEEE/CVF*计算机视觉与模式识别会议论文集, 第4183–4192页, 2020年. 10[6] Paul Bergmann, Xin Jin, David Sattlegger, a nd Carsten Ste- ger. 用于无监督3D异常检测和定位的MVTec 3D-AD数据集。 *arXivpreprint arXiv:211209045*, 2021年3..

[7] Paul Bergmann、Sindy L¨owe、Michael Fauser、David Sattlegger 和 Carsten Steger。通过将结构相似性应用于自编码器来改进无监督缺陷分割。arXiv 预印本 arXiv:1807.02011，2018年。4[8]Sangnie Bhardwaj、Willie McClinton、Tongzhou Wang、Guillaume Lajoie、Chen Sun、Phillip Isola 和 Dilip Krishnan。可转向等变表征学习。arXiv 预印本 arXiv:2302.11349，2023年。9[9] Yajie Cui、Zhaoxiang Liu 和 Shiguo Lian。无监督工业异常检测算法综述。arXiv 预印本 arXiv:2204.11161，2022年。3[10] Thomas Defard、Aleksandr Setkov、Angelique Loesch 和 Romaric Audigier。PaDiM：一种用于异常检测与定位的图像块分布建模框架。arXiv 预印本 arXiv:2011.08785，2020年。3[11] David Dehaene、Oriel Frigo、S´ebastien Combrexelle 和 Pierre Eline。在正常数据流形上进行迭代式基于能量的投影，以实现异常定位。arXiv 预印本 arXiv:2002.03734，2020年。4[12] Hanqiu Deng 和 Xingyu Li。通过从单类嵌入进行反向蒸馏实现异常检测。发表于 IEEE/CVF计算机视觉与模式识别会议论文集，第9737–9746页，2022年。1、4、6、9、10、11、13、14[13] Alexey Dosovitskiy、Lucas Beyer、Alexander Kolesnikov、Dirk Weissenborn、Xiaohua Zhai、Thomas Unterthiner、Mostafa Dehghani、Matthias Minderer、Georg Heigold、Sylvain Gelly 等。《一张图像胜过16×16个词：面向大规模图像识别的Transformer》。发表于国际学习表征会议，2020年。7、9[14]Denis Gudovskiy、Shun Ishizaka 和 Kazuki Kozuka。CFlow-AD：通过条件归一化流实现实时无监督异常检测与定位。发表于 IEEE/CVF冬季计算机视觉应用会议论文集，第98–107页，2022年。3[15] Kaiming He、Xinlei Chen、Saining Xie、Yanghao Li、Piotr Doll´ar 和 Ross Girshick。掩码自编码器是可扩展视觉学习器。发表于 IEEE/CVF计算机视觉与模式识别会议论文集，第16000–16009页，2022年。9、14[16] Kaiming He、Xiangyu Zhang、Shaoqing Ren 和 Jian Sun。用于图像识别的深度残差学习。发表于 IEEE计算机视觉与模式识别会议论文集，第 770 778, 2016 10 –页。Dan [17] Hendrycks 和 Kevin Gimpel。高斯误差线性单元（GELU）。arXiv 预印本 arXiv: *1606 08415*, 2016 8 。[18] Geoffrey Hinton、Oriol Vinyals 和 Jeff Dean。在神经网络中蒸馏知识。arXiv 预印本 arXiv:1503.02531，2015年。9[19] Jian Huang、Junzhe Wang、Yihua Tan、Dongrui Wu 和 Yu Cao。一种使用计算机视觉和巡检机器人的自动模拟仪表读数系统。《IEEE仪器与测量汇刊》，69(9)：6322–6335，2020年。1[20] Xi Jiang、Guoyang Xie、Jinbao Wang、Yong Liu、Chengjie Wang、Feng Zheng 和 Yaochu Jin。视觉传感异常检测综述。arXiv 预印本 arXiv:2202 07006，2022 3。

[21] Chun-Liang Li、Kihyuk Sohn、Jinsung Yoon 和 Tomas Pfister。CutPaste：用于异常检测与定位的自监督学习。发表于 IEEE/CVF计算机视觉与模式识别会议论文集，第9664–9674页，2021年。1、4[22] Dawei Li、Yida Li、Qian Xie、Yuxiang Wu、Zhenghao Yu 和 Jun Wang。通过由粗到细框架检测高分辨率航空发动机叶片图像中的微小缺陷。《IEEE仪器与测量汇刊》，70：1–12，2021年。5[23] Yanghao Li、Hanzi Mao、Ross Girshick 和 Kaiming He。探索用于目标检测的普通视觉Transformer骨干网络。发表于《计算机视觉—ECCV 2022：第17届欧洲会议》，以色列特拉维夫，2022年10月23日至27日，论文集，第IX部分，第280–296页。施普林格，2022年。7、8[24] Tsung-Yi Lin、Piotr Doll´ar、Ross Girshick、Kaiming He、Bharath Hariharan 和 Serge Belongie。用于目标检测的特征金字塔网络。发表于 IEEE计算机视觉与模式识别会议论文集，第2117–2125页，2017年。7[25] Ye Liu、Jun Chen 和 Jia-ao Hou。从注意力中学习位置信息：端到端弱监督裂纹分割与生成对抗网络。《工业计算机》，149：103921，2023年。3[26] Ilya Loshchilov 和 Frank Hutter。解耦权重衰减正则化。arXiv 预印本arXiv:1711.05101，2017年。10[27] Takashi Matsubara、Kazuki Sato、Kenta Hama、Ryosuke Tachibana 和 Kuniaki Uehara。使用未正则化分数进行异常检测的深度生成模型及其异构复杂性。《IEEE控制论汇刊》，52(6)：5161–5173，2020年。4[28] Subhrajit Nag、Dhruv Makwana、Sparsh Mittal、C Krishna Mohan 等。WaferSegClassNet：一种用于半导体晶圆缺陷分类与分割的轻量级网络。《工业计算机》，142：103720，2022年。1[29] Christopher Naverschnigg、Ernst Csencsics 和 Georg Schitter。用于高精度光学表面检测的柔性基于机器人的在线测量系统。《IEEE仪器与测量汇刊》，71：1–9，2022年。1[30] Jonathan Pirnay 和 Keng Chai。用于异常检测的修复Transformer。发表于《图像分析与处理—ICIAP 2022：第21届国际会议》，意大利莱切，2022年5月23日至27日，论文集，第II部分，第394–406页。施普林格，2022年。4、9、10、11、13[31] Domen Ra˘cki、Dejan Toma˘zevi˘c 和 Danijel Sko˘caj。使用卷积神经网络检测药品固体口服剂型的表面缺陷。《神经计算与应用》，34(1)：631–650，2022年。3[32] Miftahur Rahman、Haochen Liu、Mohammed Masri、Isidro Durazo-Cardenas 和 Andrew Starr。一种在移动操作机器人上使用机器人视觉进行铁路轨道重建的方法：一种拟议策略。《工业计算机》，148：103900，2023年。3[33] Karsten Roth、Latha Pemula、Joaquin Zepeda、Bernhard Sch¨olkopf、Thomas Brox 和 Peter Gehler。迈向工业异常检测中的完全召回。发表于IEEE/CVF计算机视觉与模式识别会议，页码14318–14328，2022年。1, 3, 6, 10, 11, 13, 14[34] Olga Russakovsky、Jia Deng、Hao Su、Jonathan Krause、Sanjeev Satheesh、Sean Ma、Zhiheng Huang、Andrej Karpathy、Aditya Khosla、Michael Bernstein等。ImageNet大规模视觉识别挑战赛。《国际计算机视觉杂志》，115(3):211–252，2015年。3[35]Mohammadreza Salehi、Niousha Sadjadi、Soroosh Baselizadeh、Mohammad H Rohban和Hamid R Rabiee。用于异常检测的多分辨率知识蒸馏。载于IEEE/CVF计算机视觉与模式识别会议论文集，页码14902–14912，2021年。1, 4, 9[36] Hannah M Schl¨uter、Jeremy Tan、Benjamin Hou和Bernhard Kainz。用于自监督异常检测与定位的自然合成异常。载于《计算机视觉—ECCV 2022》：第17届欧洲计算机视觉会议，特拉维夫，以色列，2022年10月23–27日，论文集，第XXXI部分，页码474–489。施普林格，2022年。1, 4, 6, 10, 11, 13[37] Xiangwen Shi、Shaobing Zhang、Miao Cheng、Lian He、Xianghong Tang和Zhe Cui。面向工业缺陷识别的小样本语义分割。《工业计算机》，148:103901，2023年。3˘[38] Domen Tabernik、Samo Sela、Jure Skvar˘c和 Danijel Sko˘caj。基于分割的深度学习表面缺陷检测方法。《智能制造期刊》，31(3):759–776，2020年。3[39] Xian Tao、Xinyi Gong、Xin Zhang、Shaohua Yan和 Chandranath Adak。工业图像无监督异常定位的深度学习：综述。IEEE仪器与测量汇刊，2022年。3[40] Matthias Wieler和Tobias Hahn。用于工业光学检测的弱监督学习。载于DAGM研讨会，2007年。3[41] Jiang Xi、Jianlin Liu、Jinbao Wang、Qiang Nie、WU Kai、Yong Liu、Chengjie Wang和Feng Zheng。Softpatch：利用噪声数据进行无监督异常检测。载于《神经信息处理系统进展》。3[42] Jie Yang、Yong Shi和Zhiquan Qi。DFR：用于无监督异常分割的深度特征重建。arXiv预印本arXiv:2012.07122，2020年。1[43] Pingping Yang、Wenhui Yue、Jian Li、Guangfu Bin和Chao Li。基于冲击特性的航空发动机叶片损伤机制与防护综述。《工程失效分析》，页码106570，2022年。6[44] Zhiyuan You、Lei Cui、Yujun Shen、Kai Yang、Xin Lu、Yu Zheng和Xinyi Le。用于多类别异常检测的统一模型。arXiv预印本arXiv:2206.03687，2022年。1, 4[45] Jiawei Yu、Ye Zheng、Xiang Wang、Wei Li、Yushuang Wu、Rui Zhao和Liwei Wu。FastFlow：通过二维正则化流实现无监督异常检测与定位。arXiv预印本arXiv:2111.07677，2021年。3[46] Sergey Zagoruyko和 Nikos Komodakis。宽残差网络。arXiv预印本arXiv:1605.07146，2016年。10

[47] Vitjan Zavrtanik、Matej Kristan和Danijel Skoˇcaj。DRAEM——一种经判别式训练、用于表面异常检测的重建嵌入。载于IEEE/CVF国际计算机视觉大会论文集，页码8330–8339，2021年。1, 4, 6, 10, 11, 13[48] Vitjan Zavrtanik、Matej Kristan和Danijel Skoˇcaj。通过图像修复实现视觉异常检测。《模式识别》，112:107706，2021年。4, 9, 10, 11, 13.[49] Alexander Zeiser、Bekir Ozcan、Bas van Stein和Thomas B¨ack。采用以数据为中心的方法评估面向在线检测的深度无监督异常检测方法。《工业计算机》，146:103852，2023年。1
[50] Xingxuan Zhang、Linjun Zhou、Renzhe Xu、Peng Cui、Zheyan Shen和Haoxin Liu。Nico++：迈向更完善的域泛化基准测试。arXiv预印本arXiv:2204.08040，2022年。3
[51]Bingchen Zhao、Shaozuo Yu、Wufei Ma、Mingxin Yu、Shenxiao Mei、Angtian Wang、Ju He、Alan Yuille和Adam Kortylewski。OOD-CV：用于评估自然图像中个体干扰因素分布偏移鲁棒性的基准。载于《计算机视觉—ECCV 2022》：第17届欧洲计算机视觉会议，特拉维夫，以色列，2022年10月23–27日，论文集，第VIII部分，页码163–180。施普林格，2022年。3[52] Yang Zou、Jongheon Jeong、Latha Pemula、Dongqing Zhang和Onkar Dabeer。用于异常检测与分割的“找不同”自监督预训练。载于《计算机视觉—ECCV 2022》：第17届欧洲计算机视觉会议，特拉维夫，以色列，2022年10月23–27日，论文集，第XXX部分，页码392–408。施普林格，2022年。1, 3