

SimpleNet：一种用于图像异常检测与定位的简单网络

刘志康¹ 周益铭² 徐元生² 王自力^{1*} 中国科学技术大学自动化系¹ 美卡
科技有限公司²

lzk@mail.ustc.edu.cn zhouyiming.donal@gmail.com xys-tc@hotmail.com zlwang@ustc.edu.cn

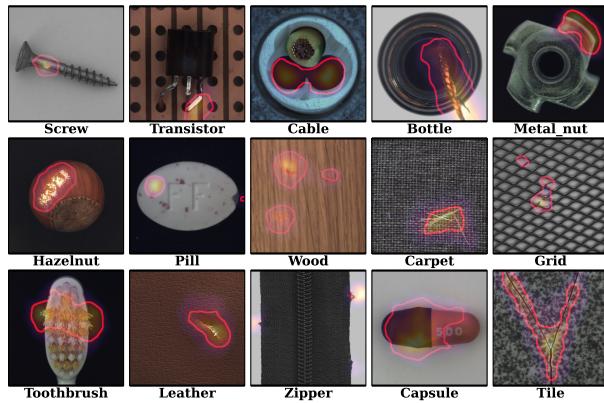


图1. MVTec AD 中样本的可视化。生成的异常图叠加在图像上。高异常分数区域以橙色标示。红色边界表示实际异常分割图的轮廓。

摘要

We propose a simple and application-friendly network (called SimpleNet) for detecting and localizing anomalies. SimpleNet consists of four components: (1) a pre-trained Feature Extractor that generates local features, (2) a shallow Feature Adapter that transfers local features towards target domain, (3) a simple Anomaly Feature Generator that counterfeits anomaly features by adding Gaussian noise to normal features, and (4) a binary Anomaly Discriminator that distinguishes anomaly features from normal features. During inference, the Anomaly Feature Generator would be discarded. Our approach is based on three intuitions. First, transforming pre-trained features to target-oriented features helps avoid domain bias. Second, generating synthetic anomalies in feature space is more effective, as defects may not have much commonality in the image space. Third, a simple discriminator is much efficient and practical. In spite of simplicity, SimpleNet outperforms previous methods quantitatively and qualitatively. On

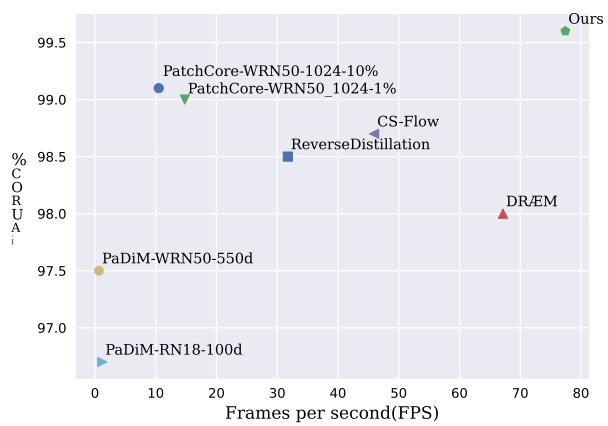


图2. MVTec AD基准测试中推理速度 (FPS) 与I-AUROC的关系。SimpleNet在准确性和效率上均大幅超越所有先前方法。

the MVTec AD benchmark, SimpleNet achieves an anomaly detection AUROC of 99.6%, reducing the error by 55.5% compared to the next best performing model. Furthermore, SimpleNet is faster than existing methods, with a high frame rate of 77 FPS on a 3080ti GPU. Additionally, SimpleNet demonstrates significant improvements in performance on the One-Class Novelty Detection task. Code: <https://github.com/DonaldRR/SimpleNet>.

1. 引言

图像异常检测与定位任务旨在识别异常图像并定位异常子区域。检测各类目标异常的技术在工业检测领域具有广泛的应用[3, 6]。在工业场景中，异常检测与定位尤为困难，因为异常样本稀缺且异常形态多变——从细微变化（如细划痕）到大型结构缺陷e.g乃至零件缺失皆有可能。图1展示了MVTec AD基准数据集[3]的部分样本及我们提出方法的检测结果。此类情况

*Corresponding author

限制了监督方法向 $\{v^*\}$ 靠近。

当前方法以无监督方式处理此问题，训练过程中仅使用正常样本。基于重建的方法[10, 21, 31]、基于合成的方法[17, 30]以及基于嵌入的方法[6, 22, 24]是解决该问题的三大主流方向。基于重建的方法（如[21, 31]）假设仅用正常数据训练的深度网络无法准确重建异常区域，并将像素级重建误差作为异常定位的评分依据。然而这一假设并非始终成立，网络有时可能“泛化”能力过强，以致对异常输入也能进行良好重建，从而导致漏检[10, 19]。基于合成的方法[17, 30]通过在无异常图像上生成合成异常进行训练，以估计正常与异常之间的决策边界。但合成图像的真实性不足，其提取的特征可能严重偏离正常特征分布，使用此类负样本训练可能导致正常特征空间的边界过于宽松，致使不明显的缺陷被纳入正常特征空间。

最近，基于嵌入的方法[6, 7, 22, 24]实现了最先进的性能。这些方法使用ImageNet预训练的卷积神经网络（CNN）来提取广义的正常特征。随后采用多元高斯分布[6]、标准化流[24]和记忆库[22]等统计算法来嵌入正常特征的分布。通过将输入特征与学习到的分布或记忆的特征进行比较来检测异常。然而，工业图像的分布通常与ImageNet不同。直接使用这些有偏差的特征可能导致不匹配问题。此外，统计算法往往面临高计算复杂度或高内存消耗的问题。

为缓解上述问题，我们提出了一种名为SimpleNet的新型异常检测与定位网络。SimpleNet融合了基于合成和基于嵌入的方法思路，并进行了多项改进。首先，我们提出使用特征适配器生成面向目标的特征，而非直接使用预训练特征，从而减少领域偏差。其次，我们提出通过在特征空间中对正常特征施加噪声来生成异常特征，而非直接在图像上合成异常。我们认为通过合理校准噪声尺度，可以获得紧密有界的正常特征空间。第三，我们通过训练一个简单的判别器来简化异常检测流程，这比前述基于嵌入方法采用的复杂统计算法在计算效率上更具优势。具体而言，SimpleNet利用预训练主干网络提取正常特征，随后通过特征适配器

将特征转移到目标域中。然后，通过向适应后的正常特征添加高斯噪声，简单地生成异常特征。一个由几层MLP组成的简单判别器在这些特征上进行训练，以区分异常。

SimpleNet易于训练和应用，具有出色的性能和推理速度。所提出的SimpleNet基于广泛使用的WideResnet5.0主干网络，在MVTec AD数据集上实现了99.6%的AUROC，同时以77帧/秒的速度运行，在准确性和效率上均超越了先前已发表的最佳异常检测方法，如图2所示。我们进一步将SimpleNet引入单类新颖性检测任务，以展示其通用性。这些优势使SimpleNet能够弥合学术研究与工业应用之间的差距。代码将公开提供。

2. 相关工作

异常检测与定位方法主要可分为三类：*i.e.*，基于重构的方法、基于合成的方法和基于嵌入的方法。

基于重建的方法秉持这样的观点：异常图像区域不应被正确重建，因为它们不存在于训练样本中。一些方法[10]利用生成模型，如自编码器和生成对抗网络[11]，对正常数据进行编码和重建。其他方法[13, 21, 31]将异常检测视为图像修复问题，随机掩码图像中的局部块，然后利用神经网络预测被擦除的信息。在训练中广泛采用结构相似性指数（SSIM）[29]损失函数进行整合。异常图通过输入图像与其重建图像之间的像素级差异生成。然而，如果异常与正常训练数据共享常见的结构模式（例如局部边缘），或者解码器“过于强大”以至于能很好地解码某些异常编码，则图像中的异常很可能被较好地重建[31]。

基于合成的方法通常在无异常图像上合成异常。DRÆM [30] 提出了一种网络，该网络以端到端的方式在合成生成的刚好超出分布模式的图像上进行判别性训练。CutPaste [17] 提出了一种简单的策略来生成用于异常检测的合成异常，即切割一个图像块并将其粘贴到大图像的随机位置。训练一个CNN来区分来自正常和增强数据分布的图像。然而，合成异常的外观与真实异常并不完全匹配。实际上，由于缺陷多种多样且不可预测，生成一个包含所有异常值的异常集是不可能的。与在图像上合成异常不同，通过提出的SimpleNet，负样本

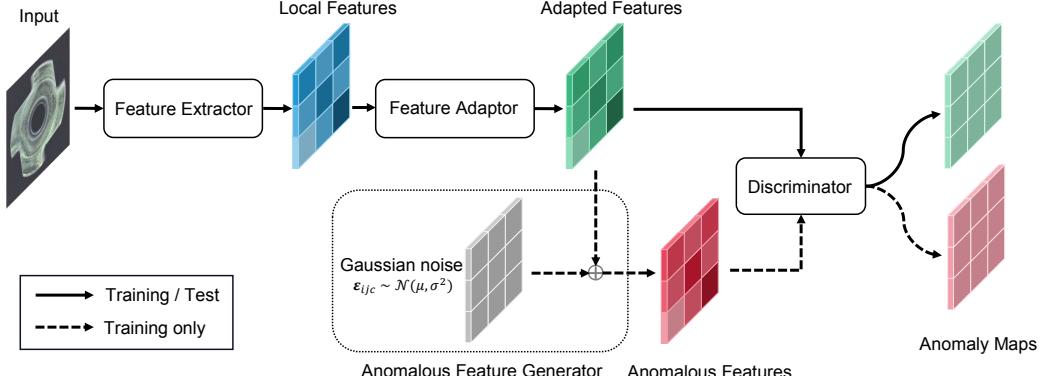


图3. 所提出的SimpleNet概述。在训练阶段，将正常样本输入预训练的Feature Extractor以获取局部特征。随后，利用Feature Adaptor将预训练特征适配到目标域。通过向适配特征添加高斯噪声来合成异常特征。适配特征与异常特征分别作为正负样本，用于训练最终的Discriminator。在推理阶段，Anomalous Feature Generator被移除。

在特征空间中合成。

基于嵌入的方法近年来实现了最先进的性能。这些方法将正常特征嵌入到压缩空间中，异常特征在嵌入空间中远离正常聚类。典型方法[6,7,22,24]利用在ImageNet上预训练的网络进行特征提取。通过预训练模型，PaDiM[6]将提取的异常补丁特征通过多元高斯分布进行嵌入。PatchCore[22]则使用最大代表性的正常补丁特征记忆库。在测试中采用马氏距离或最大特征距离对输入特征进行评分。然而，工业图像通常具有与ImageNet不同的分布，直接使用预训练特征可能导致不匹配问题。此外，无论是计算协方差矩阵的逆[6]还是在记忆库中搜索最近邻[22]，都会限制实时性能，特别是在边缘设备上。

CS-Flow [24]、CFLOW-AD [12] 和 DifferNet [23] 提出通过归一化流 (NF) [20] 将正常特征分布转换为高斯分布。由于归一化流只能处理全尺寸特征图（即不允许下采样），且耦合层 [9] 的内存消耗是普通卷积层的数倍，这些方法较为消耗内存。蒸馏方法 [4, 7] 训练一个学生网络，使其仅通过正常样本匹配固定预训练教师网络的输出。当输入异常查询时，应能检测到学生与教师输出之间的差异。由于输入图像需同时经过教师和学生网络，计算复杂度增加了一倍。

SimpleNet克服了上述问题。它采用了一个特征适配器，在目标数据集上进行迁移学习，以减轻预训练CNN的偏差。SimpleNet提出通过合成异常样本来实现这一目标。

SimpleNet在推理时采用单流方式，完全由传统CNN模块构建，这有助于快速训练、推理及工业应用，其操作聚焦于特征空间而非直接作用于图像。

3. 方法

本节详细介绍了所提出的SimpleNet。如图3所示，SimpleNet由一个Feature Extractor、一个Feature Adaptor、一个Anomalous Feature Generator和一个Discriminator组成。Anomalous Feature Generator仅在训练期间使用，因此SimpleNet在推理时遵循单流模式。接下来将依次描述这些模块。

3.1. 特征提取器

特征提取器按照[22]的方式获取局部特征。我们将该过程重新表述如下：将训练集和测试集分别记为 \mathcal{X}_{train} 和 \mathcal{X}_{test} 。对于 $\mathcal{X}_{train} \cup \mathcal{X}_{test}$ 中的任意图像 $x_i \in \mathbb{R}^{H \times W \times 3}$ ，预训练网络 ϕ 会像ResNet类主干网络通常所做的那样，从不同层级提取特征。由于预训练网络会偏向其训练所用的数据集，因此仅选择部分层级用于目标数据集是合理的。形式上，我们定义 L 为包含所用层级索引的子集。来自层级 $l \in L$ 的特征图记为 $\phi^{l,i} \sim \phi^l(x_i) \in \mathbb{R}^{H_l \times W_l \times C_l}$ ，其中 H_l 、 W_l 和 C_l 分别表示特征图的高度、宽度和通道数。对于位置 (h, w) 处的特征单元 $\phi_{h,w}^{l,i} \in \mathbb{R}^{C_l}$ ，其邻域（补丁尺寸为 p ）定义为

$$\mathcal{N}_p^{(h,w)} = \{(h', w') | h' \in [h - \lfloor p/2 \rfloor, \dots, h + \lfloor p/2 \rfloor], w' \in [w - \lfloor p/2 \rfloor, \dots, w + \lfloor p/2 \rfloor]\} \quad (1)$$

聚合邻域 $\mathcal{N}_p^{(h,w)}$ 内的特征，使用聚合函数 f_{agg} （采用自适应平均池化

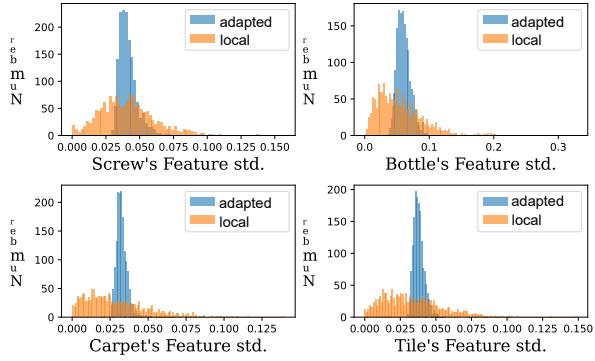


图4. 局部特征与适配特征各维度标准差直方图。在异常特征训练时，适配特征空间变得更加紧凑。

在此处进行)得到局部特征 $z_{h,w}^{l,i}$, 如

$$z_{h,w}^{l,i} = f_{agg}(\{\phi_{h',y'}^{l,i}|(h',y') \in \mathcal{N}_p^{h,w}\}) \quad (2)$$

为了融合来自不同层次的特征 $z_{h,w}^{l,i}$, 所有特征图均被线性调整至相同尺寸 (H_0, W_0) , 即最大特征图的尺寸。通过简单地沿通道维度拼接特征图, 即可得到特征图 $o^i \in \mathbb{R}^{H_0 \times H_0 \times C}$ 。该过程定义为

$$o^i = f_{cat}(resize(z^{l',i}, (H_0, W_0))|l' \in L \quad (3)$$

我们将 $o_{h,w}^i \in \mathbb{R}^C$ 定义为 o^i 在位置 (h, w) 处的条目。我们将上述表达式简化为

$$o^i = F_\phi(x^i) \quad (4)$$

其中 F_ϕ 是特征提取器。

3.2. 特征适配器

由于工业图像通常与骨干网络预训练所使用的数据集具有不同的分布, 我们采用特征适配器 G_θ 将训练特征迁移到目标域。该特征适配器 G_θ 将局部特征 $o_{h,w}$ 映射为适应后的特征 $q_{h,w}$, 具体表示为:

$$q_{h,w}^i = G_\theta(o_{h,w}^i) \quad (5)$$

特征适配器可以由简单的神经块组成, 例如全连接层或多层感知机 (MLP)。我们通过实验发现, 单个全连接层即可获得良好的性能。

3.3. 异常特征生成器

为了训练判别器估计样本正常的可能性, 最简单的方法是采样负样本, 即缺陷特征, 并将其与正常样本一起优化。缺陷的缺乏使得采样

分布估计难以处理。尽管[17, 18, 30]依赖额外数据合成缺陷图像, 我们在正常样本的特征空间中添加简单噪声, 并声称这种方法优于那些经过复杂处理的方法。

异常特征是通过在正常特征 $q_{h,w}^i \in \mathbb{R}^C$ 上添加高斯噪声生成的。形式上, 采样一个噪声向量 $\epsilon \in \mathbb{R}^C$, 其中每个条目遵循独立同分布的高斯分布 $\mathcal{N}(\mu, \sigma^2)$ 。异常特征 $q_{h,w}^{i-}$ 的融合方式为

$$q_{h,w}^{i-} = q_{h,w}^i + \epsilon \quad (6)$$

图4展示了异常特征对MVTec AD四种类别的影响。我们可以看到, 适应特征各维度的标准差趋于一致。因此, 在区分异常特征与正常特征时, 特征空间倾向于变得紧凑。

3.4. 判别器

判别器 D_ψ 作为一个正常性评分器工作, 直接估计每个位置 (h, w) 的正常性。由于负样本与正常特征 $\{q^i|x^i \in \mathcal{X}_{train}\}$ 一同生成, 在训练过程中它们都被输入到判别器。判别器期望对正常特征输出正值, 而对异常特征输出负值。我们简单地采用常见的两层多层感知机 (MLP) 结构作为分类器, 将正常性估计为 $D_\psi(q_{h,w}) \in \mathbb{R}$ 。

3.5. 损失函数与训练

一个简单的截断 $l1$ 损失被推导为

$$l_{h,w}^i = \max(0, th^+ - D_\psi(q_{h,w}^i)) + \max(0, -th^- + D_\psi(q_{h,w}^{i-})) \quad (7)$$

th^+ 和 th^- 是防止过拟合的截断项, 默认设置为 0.5 和 -0.5。训练目标为

$$\mathcal{L} = \min_{\theta, \psi} \sum_{x^i \in \mathcal{X}_{train}} \sum_{h,w} \frac{l_{h,w}^i}{H_0 * W_0} \quad (8)$$

我们将在实验部分通过实验评估所提出的截断 $\{v^*\}1$ 损失函数与广泛使用的交叉熵损失。训练过程的伪代码如算法1所示。

3.6. 推理与评分函数

异常特征生成器在推理阶段被弃用。需要注意的是, 其余模块可以堆叠成一个端到端网络。我们将每个 $x_i \in \mathcal{X}_{test}$ 依次输入前述特征提取器 F_ϕ 和特征适配器 G_θ , 以获取如公式5所示的适配特征 $q_{h,w}^i$ 。异常分数由判别器提供。

D_ψ as

$$s_{h,w}^i = -D_\psi(q_{h,w}^i) \quad (9)$$

推理过程中用于异常定位的异常图定义为

$$S_{AL}(x_i) := \{s_{h,w}^i | (h, w) \in W_0 \times H_0\} \quad (10)$$

随后，将 $S_{AL}(x_i)$ 插值至输入样本的空间分辨率，并使用 $\sigma = 4$ 进行高斯滤波以获得平滑边界。由于异常区域无论大小均存在最敏感响应点，因此取异常图的最大得分作为每张图像的异常检测分数。

$$S_{AD}(x_i) := \max_{(h,w) \in W_0 \times H_0} s_{h,w}^i \quad (11)$$

4. 实验

4.1. 数据集。

我们在MVTec异常检测基准[3]上进行了大部分实验，该数据集是异常检测与定位领域的知名数据集。MVTec AD包含5种纹理类别和10种物体类别，均源自工业制造场景，共计5354张图像。数据集由用于训练的正常图像，以及同时包含正常图像与多种缺陷类型的异常测试图像构成，并为缺陷测试图像提供了像素级标注。典型图像示例如图1所示。参照[6, 22]的方法，我们将图像分别调整尺寸并中心裁剪至 256×256 和 224×224 。实验过程中未采用数据增强技术。我们遵循单分类协议（亦称冷启动异常检测），即为每个类别分别使用其正常训练样本训练独立的单分类器。

我们在CIFAR10数据集[16]上进行了单类新颖性检测，该数据集包含5万张训练图像和1万张测试图像，图像尺寸为 32×32 ，共分为10个类别。在单类新颖性检测的设置下，其中一个类别被视为正常数据，其余类别则作为新颖性数据。

4.2. 评估指标。

图像级异常检测性能通过标准的接收者操作特征曲线下面积（我们记为I-AUROC）来衡量，使用生成的异常检测分数 S_{AD} （公式11）。对于异常定位，异常图 S_{AL} （公式10）用于评估像素级AUROC（记为P-AUROC）。根据先前的研究[6, 22]，我们在MVTec AD上计算类别平均AUROC以及检测和定位在所有类别上的平均AUROC。比较基线包括AE-SSIM [3]、RIAD [31]、DRÆM [30]、CutPaste [17]、CS-Flow [24]、PaDiM [6]、RevDist [7] 和 PatchCore [22]。

Algorithm 1 SimpleNet training pseudo-code, Pytorch-like

```

# F: Feature Extractor
# G: Feature Adaptor
# N: i.i.d Gaussian noise
# D: Discriminator
pretrain_init(F)
random_init(G, D)
for x in data_loader:
    o = F(x) # normal features
    q = G(o) # adapted features
    q_ = q + random(N) # anomalous features

    loss = loss_func(D(q), D(q_)).mean()
    loss.backward() # back-propagate

    F = F.detach() # stop gradient
    update(G, D) # Adam

# loss function
def loss_func(s, s_):
    th_ = -th = 0.5
    return max(0, th-s) + max(0, th+s_)

```

4.3. 实现细节

本节阐述了本文实验的配置实现细节。实验中使用的所有骨干网络均在ImageNet[8]上进行了预训练。当骨干网络采用类ResNet架构时，特征提取器遵循[22]的方法，使用骨干网络的第2和第3中间层（如公式3中的 $l' \in [2,3]$ ）。默认情况下，我们的实现采用WideResnet 50作为骨干网络，特征提取器输出的特征维度设置为1536。后续的特征适配器本质上是一个无偏置的全连接层，该适配器中FC层的输入与输出特征维度相同。异常特征生成器对正常特征的每个条目添加独立同分布的高斯噪声 $\mathcal{N}(0, \sigma^2)$ ，默认设定 $\sigma=0.015$ 。后续判别器由线性层、批归一化层、Leaky ReLU（斜率0.2）和线性层构成。公式7中的 th^+ 和 th^- 均设置为0.5。采用Adam优化器，将特征适配器和判别器的学习率分别设为0.0001和0.0002，权重衰减设为0.00001。每个数据集的训练轮数设置为160，批大小为4。

4.4. MVTec AD数据集上的异常检测

MVTec AD上的异常检测结果如表1所示。图像级异常分数由异常图的最大分数给出，如公式11所示。SimpleNet在15个类别中的9个类别上取得了最高分数。对于纹理和物体，SimpleNet分别达到了99.8%和99.5%的I-AUROC新SOTA。SimpleNet实现了显著更高的平均图像异常检测性能，即99.6%的I-AUROC分数。请注意

表1. SimpleNet与MVTec AD上最先进工作的比较。每个条目中展示了图像级AUROC (I-AUROC) 和像素级AUROC (P-AUROC)，格式为I-AUROC%/P-AUROC%。CS-Flow的P-AUROC在文献[24]中未记录。

Type	Reconstruction-based		Synthesizing-based		Embedding-based				Ours
Model	AE-SSIM	RIAD	DRÆM	CutPaste	CS-Flow	PaDiM	RevDist	PatchCore	SimpleNet
Carpet	87/64.7	84.2/96.3	97.0/95.5	93.9/98.3	100/-	99.8/99.1	98.9/98.9	98.7/99.0	99.7/98.2
Grid	94/84.9	99.6/98.8	99.9/99.7	100/97.5	99.0/-	96.7/97.3	100/99.3	98.2/98.7	99.7/98.8
Leather	78/56.1	100/99.4	100/98.6	100/99.5	100/-	100/99.2	100/99.4	100/99.3	100/99.2
Tile	59/17.5	98.7/89.1	99.6/99.2	94.6/90.5	100/-	98.1/94.1	99.3/95.6	98.7/95.6	99.8/97.0
Wood	73/60.3	93.0/85.8	99.1/96.4	99.1/95.5	100/-	99.2/94.9	99.2/95.3	99.2/95.0	100/94.5
Avg. Text.	78/56.7	95.1/93.9	99.1/97.9	97.5/96.3	99.8/-	95.5/96.9	99.5/97.7	99.0/97.5	99.8/97.5
Bottle	93/83.4	99.9/98.4	99.2/99.1	98.2/97.6	99.8/-	99.1/98.3	100/98.7	100/98.6	100/98.0
Cable	82/47.8	81.9/84.2	91.8/94.7	81.2/90.0	99.1/-	97.1/96.7	95.0/97.4	99.5/98.4	99.9/97.6
Capsule	94/86.0	88.4/92.8	98.5/94.3	98.2/97.4	97.1/-	87.5/98.5	96.3/98.7	98.1/98.8	97.7/98.9
Hazelhut	97/91.6	83.3/96.1	100/99.7	98.3/97.3	99.6/-	99.4/98.2	99.9/98.9	100/98.7	100/97.9
Metal Nut	89/60.3	88.5/92.5	98.7/99.5	99.9/93.1	99.1/-	96.2/97.2	100/97.3	100/98.4	100/98.8
Pill	91/83.0	83.8/95.7	98.9/97.6	94.9/95.7	98.6/-	90.1/95.7	96.6/98.2	96.6/97.4	99.0/98.6
Screw	96/88.7	84.5/98.8	93.9/97.6	88.7/96.7	97.6/-	97.5/98.5	97.0/ 99.6	98.1/99.4	98.2/99.3
Toothbrush	92/78.4	100/98.9	100/98.1	99.4/98.1	91.9/-	100/98.8	99.5/ 99.1	100/98.7	99.7/98.5
Transistor	90/72.5	90.9/87.7	93.1/90.9	96.1/93.0	99.3/-	94.4/97.5	96.7/92.5	100/96.3	100/97.6
Zipper	88/66.5	98.1/97.8	100/98.8	99.9/99.3	99.7/-	98.6/98.5	98.5/98.2	99.4/98.8	99.9/98.9
Avg. Obj.	91/75.8	89.9/94.3	97.4/97.0	95.5/95.8	98.2/-	96.0/97.8	98/97.9	99.2/98.4	99.5/98.4
Average	87/69.4	91.7/94.2	98.0/97.3	96.1/96.0	98.7/-	95.8/97.5	98.5/97.8	99.1/98.1	99.6/98.1

表2. 在不同WideResNet50层级组合下， MVTec AD的性能表现。

level1	level2	level3	I-AUROC%	P-AUROC%
✓	✓	✓	93.0	94.2
			98.4	96.7
	✓	✓	99.2	97.5
	✓	✓	96.7	96.7
✓	✓	✓	99.6	98.1
✓	✓	✓	99.1	98.1

从PatchCore [22] (同等WideResnet50骨干网络下的次优竞争者) 0.9%的误差降至SimpleNet的0.4%，意味着误差减少了55.5%。在工业检测场景中，这是相关且显著的降低。

4.5. MVTec AD数据集上的异常定位

异常定位性能通过像素级AUROC来衡量，我们将其记为P-AUROC。与最先进方法的比较如表1所示。SimpleNet在MVTec AD上实现了98.1% P-AUROC的最佳异常检测性能，同时在物体类别上以98.4% P-AUROC创造了新的SOTA记录。在15个类别中，SimpleNet在4个类别上取得了最高分。我们在图8中展示了异常定位的代表性样本可视化结果。

4.6. 推理时间

除了检测和定位性能外，推理时间是工业模型部署中最受关注的问题。图2展示了与最先进方法在推理时间上的对比。所有方法均在相同硬件条件下进行测量

使用一块Nvidia GeForce GTX 3080ti GPU和一颗Intel(R) Xeon(R) CPU E5-2680 v3@2.5GHZ处理器。结果清晰表明，我们的方法在取得最佳性能的同时也实现了最快的速度。SimpleNet比PatchCore [22]快了近8×倍。

4.7. 消融研究

邻域大小与层级结构。我们研究了公式1中邻域大小 p 的影响。图6结果显示异常预测在局部性与全局上下文之间存在明确的最优平衡，因此我们选择邻域大小 $p = 3$ 。我们设计了一组实验来测试层级子集 L 对模型性能的影响，结果如表2所示。我们将前三个WideResNet50模块编号为1–3。可见，仅使用第3层级特征已能达到最优性能，但结合第2层级特征可进一步提升效果。我们最终选择2+3作为默认设置。

适配器配置。适配器对预训练特征提供一种变换（投影）。我们默认的特征适配器是一个不带偏置的单层全连接层，输入和输出通道数相等。表3展示了不同特征适配器的对比，首行“Ours”的实现遵循与表1相同的配置。“Ours-complex-FA”将简单特征适配器替换为非线性版本（即带激活函数的单层MLP）。行“Ours-w/o-FA”则移除了特征适配器。结果表明单层全连接层能取得最佳性能。直观上，特征适配器通过寻找一种投影，使得伪造的异常特征与投影后的预训练特征易于区分，这意味着判别器只需一个简单的解决方案。这一现象也表明，使用特征适配器有助于网络收敛

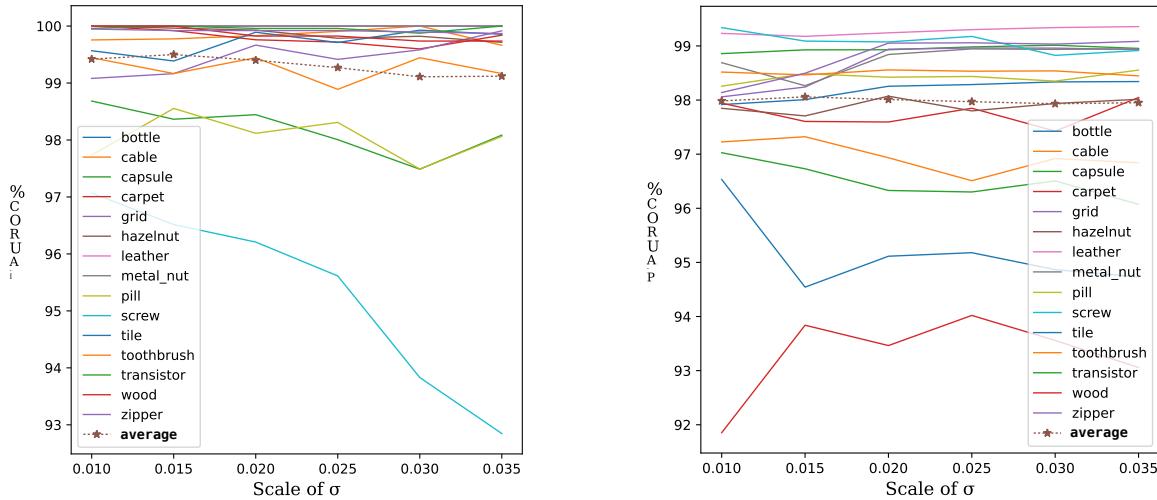


图5. 针对MVTec AD数据集中每个类别，在不同 σ 下的I-AUROC%和P-AUROC%。（建议彩色查看。）

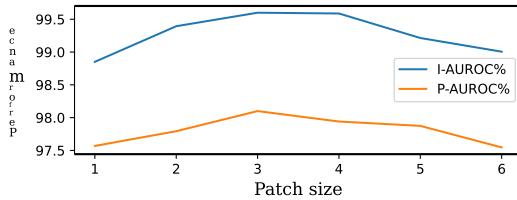


图6. 在MVTec AD数据集上不同补丁尺寸的性能表现。

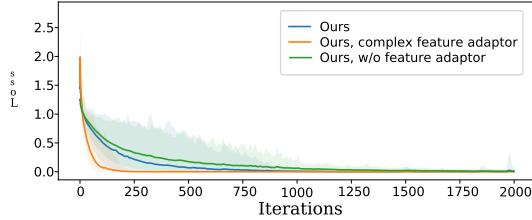


图7. 训练过程中的损失可视化。图中曲线展示了MVTec AD数据集中所有类别的平均损失。半透明色带显示了损失波动的范围。

边缘快速（图7）。我们观察到使用复杂的特征适配器时性能显著下降。一个可能的原因是复杂的适配器可能导致过拟合，降低了对测试中各种缺陷的泛化能力。图4比较了特征适配器前后各维度特征标准差的直方图。可以看出，在使用异常特征进行训练时，适配后的特征空间变得紧凑。

噪声尺度。异常特征生成器中的噪声尺度控制着合成的异常特征与正常特征之间的距离。具体来说，较高的 σ 会导致异常特征保持较高的欧几里得距离——

表3. 不同特征适配器的比较。“Ours”实现遵循与表1相同的配置。“Ours-complex-FA”将简单特征适配器替换为非线性适配器。“Ours-w/o-FA”移除了特征适配器，相当于使用恒等全连接层。“Ours-CE”使用交叉熵损失。表中展示了MVTec AD的I-AUROC%和P-AUROC%。

Model	I-AUROC%	P-AUROC%
Ours	99.6	98.1
Ours-complex-FA	98.3	97.2
Ours-w/o-FA	99.2	97.9
Ours-CE	99.4	97.8

表4. 不同骨干网络在MVTec AD上的性能表现。

Model	I-AUROC%	P-AUROC%
ResNet18	98.3	95.7
ResNet50	99.6	98.0
ResNet101	99.2	97.6
WideResNet50	99.6	98.1

表5. CIFAR-10数据集上单类新颖性检测的I-AUROC(%)结果。

Method	LSA	DSVDD	OCGAN	HRN	DAAD
AUROC	64.1	64.8	65.6	71.3	75.3
Method	DisAug CLR	IGD	MKD	RevDist	SimpleNet
AUROC	80.0	83.68	84.5	86.5	86.5

对正常特征的容忍度。在较大的 σ 上进行训练会导致决策边界宽松，从而产生较高的假阴性。反之，如果 σ 过小，训练过程将变得不稳定，且判别器无法很好地泛化到正常特征。图5详细展示了MVTec AD中每个类别 σ 的影响。可以看出， $\sigma = 0.015$ 达到了平衡并取得了最佳性能。

损失函数。我们比较了所提出的损失函数

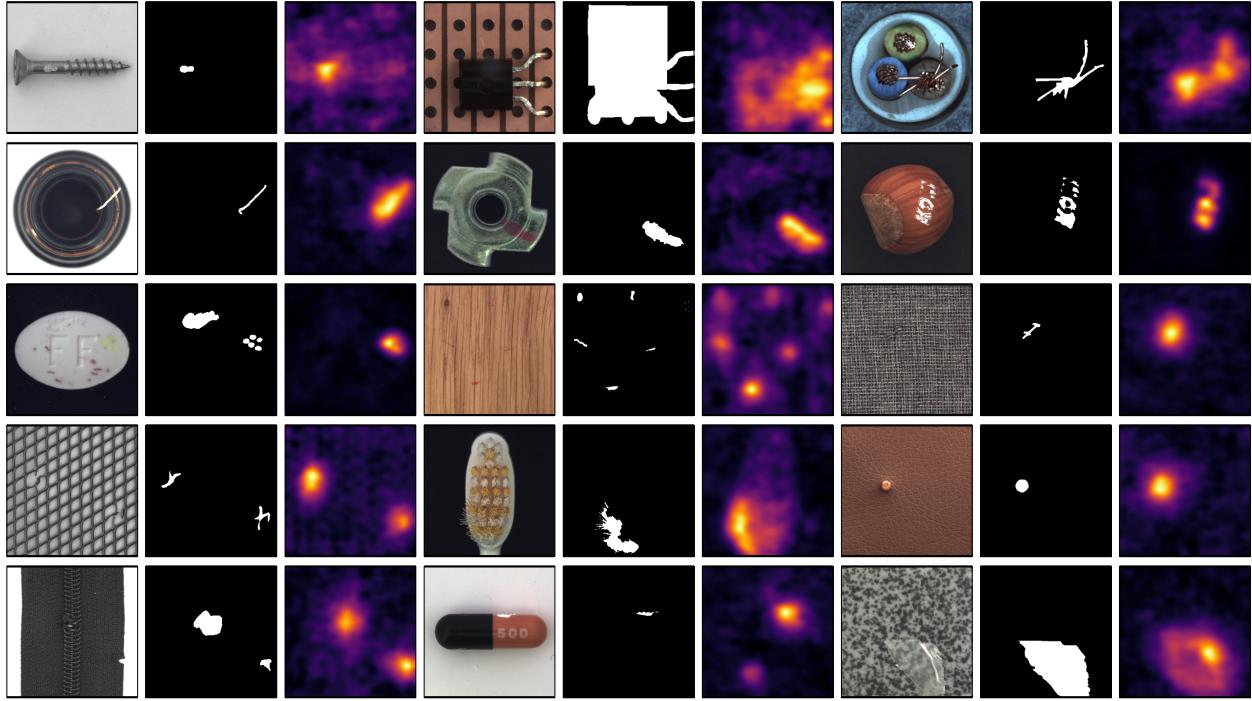


图8. 定性结果，其中为MVTec AD中的每个类别展示了采样图像、真实标注和异常图。

在3.5节中，我们使用了广泛采用的交叉熵损失（如表3中“Ours-CE”行所示）。我们发现相较于交叉熵损失，该方法在I-AUROC上提升了0.2%，在P-AUROC上提升了0.3%。

对骨干网络的依赖。我们使用不同的骨干网络测试SimpleNet，结果如表4所示。我们发现，在不同骨干网络的选择上，结果基本保持稳定。选择WideResNet50是为了与PaDiM [6] 和 PatchCore [22] 具有可比性。

定性结果图8展示了异常定位的结果，这些结果标示出了异常区域。分割结果的阈值是通过计算每个子类所有异常分数的F1分数来获得的。实验结果证明，即使在一些相当困难的情况下，所提出的方法也能很好地定位异常区域。此外，我们可以发现，所提出的方法在物体类和纹理类上均表现出一致的性能。

4.8. 单类新颖性检测

为了评估所提出的SimpleNet的泛化能力，我们在CIFAR-10数据集[16]上进行了单类新颖性检测实验。参照[19]的方法，我们使用单一类别的样本训练模型，并检测来自其他类别的新颖样本。我们分别为每个类别训练了相应的模型。请注意，新颖性分数定义为相似度图中的最高分数。表5报告了我们的方法与其他方法的I-AUROC分数。为公平比较，所有方法均在ImageNet上进行了预训练。

基线方法包括VAE [2]、LSA [1]、DSVDD [25]、OCGAN [19]、HRN [15]、AnoGAN [27]、DAAD [14]、MKD [26]、DisAug CLR [28]、IGD [5]和RevDist [7]。我们的方法优于这些对比方法。需要注意的是，当结合自监督学习进行增强时，IGD [5]和DisAug CLR [28]分别达到了91.25%和92.4%的性能。

5. 结论

本文提出了一种名为SimpleNet的简单而高效的无监督异常检测与定位方法。SimpleNet由多个易于训练且适用于工业场景的简单神经网络模块组成。尽管结构简洁，SimpleNet在MVTec AD基准测试中相比现有最优方法，不仅取得了最高性能，同时实现了最快的推理速度。该方法为弥合异常检测与定位领域学术研究与工业应用之间的鸿沟提供了新视角。

致谢

本研究得到国家自然科学基金（项目编号：62176246和61836008）的资助。同时，本研究亦获得安徽省自然科学基金（项目编号：2208085UD17）以及中央高校基本科研业务费专项资金（项目编号：WK3490000006）的支持。

参考文献

[1] Davide Abati, Angelo Porrello, Simone Calderara, 和 Rita Cucchiara。用于新颖性检测的潜在空间自回归。载于 *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 第481–490页, 2019年。

8[2] Jinwon An 和 Sungzoon Cho。基于重构概率的变分自编码器异常检测。*Special Lecture on IE*, 2(1):1–18, 2015年。8

[3] Paul Bergmann, Michael Fauser, David Sattlegger, 和 Carsten Steger。MVTec AD——一个用于无监督异常检测的综合真实世界数据集。载于 *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 第9592–9600页, 2019年。1, 5[4] Paul Bergmann, Michael Fauser, David Sattlegger, 和 Carsten Steger。不知情的学生：具有判别性潜在嵌入的师生异常检测。载于 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 第4183–4192页, 2020年。3[5] Yuanhong Chen, Yu Tian, Guansong Pang, 和 Gustavo Carneiro。通过插值高斯描述符进行深度单类分类。载于 *Proceedings of the AAAI Conference on Artificial Intelligence*, 第383–392页, 2022年。8[6] Thomas Defard, Aleksandr Setkov, Angelique Loesch, 和 Romaric Audigier。PaDiM：一种用于异常检测和定位的补丁分布建模框架。载于 *International Conference on Pattern Recognition*, 第475–489页。Springer, 2021年。1, 2, 3, 5, 8[7] Hanqiu Deng 和 Xingyu Li。通过从单类嵌入进行反向蒸馏的异常检测。载于 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022年。2, 3, 5, 8[8] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, 和 Li Fei-Fei。ImageNet：一个大规模分层图像数据库。载于 *2009 IEEE conference on computer vision and pattern recognition*, 第248–255页。IEEE, 2009年。5[9] Laurent Dinh, Jascha Sohl-Dickstein, 和 Samy Bengio。使用Real NVP进行密度估计。*arXiv preprint arXiv:1605.08803*, 2016年。3[10] Dong Gong, Lingqiao Liu, Vuong Le, Budhaditya Saha, Moussa Reda Mansour, Svetha Venkatesh, 和 Anton van den Heijzel。记忆正常以检测异常：用于无监督异常检测的记忆增强深度自编码器。载于 *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 第1705–1714页, 2019年。2[11] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, 和 Yoshua Bengio。生成对抗网络。*Advances in neural information processing systems*, 27, 2014年。2[12] Denis Gudovskiy, Shun Ishizaka, 和 Kazuki Kozuka。CFLOW-AD：通过条件归一化流进行定位的实时无监督异常检测。载于 *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 第98–107页, 2022年。3[13] Matthias Haselmann, Dieter P Gruber, 和 ul Tabatabai。使用基于深度学习的图像进

pletion. 于 *2018 17th IEEE international conference on machine learning and applications (ICMLA)*, 第1237–1242页。IEE, 2018年。2 [14] 侯金磊, 张莹莹, 钟乔勇, 谢迪, 蒲世良, 周宏。分而治之：学习块级记忆用于无监督异常检测。于 *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 第8791–8800页, 2021年。8 [15] 胡文鹏, 王梦宇, 秦琦, 马金文, 刘冰。HRN：一种整体性单类学习方法。*Advances in Neural Information Processing Systems*, 33:19111–19124, 2020年。8 [16] A. Krizhevsky。从微小图像中学习多层特征。*Master's thesis, University of Tront*, 2009年。5, 8 [17] 李春亮, 孙基赫, 尹珍成, Tomas Pfister。CutPaste：用于异常检测与定位的自监督学习。于 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 第9664–9674页, 2021年。2, 4, 5 [18] Philipp Liznerski, Lukas Ruff, Robert A. Vandermeulen, Billy Joe Franks, Marius Kloft, 和 Klaus Robert Müller。可解释的深度单类分类。于 *International Conference on Learning Representations*, 2020年。4 [19] Pramuditha Perera, Ramesh Nallapati, 向兵。OC-GAN：使用具有约束潜在表示的GAN进行单类新颖性检测。于 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 第2898–2906页, 2019年。2, 8 [20] Danilo Rezende, Shakir Mohamed。使用标准化流进行变分推断。于 *International conference on machine learning*, 第1530–1538页。PMLR, 2015年。3 [21] Nicolae-Cătălin Ristea, Neelu Madan, Radu Tudor Ionescu, Kamal Nasrollahi, Fahad Shahbaz Khan, Thomas B. Moeslund, Mubarak Shah。用于异常检测的自监督预测卷积注意力块。于 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 第13576–13586页, 2022年。2 [22] Karsten Roth, Latha Pemula, Joaquin Zepeda, Bernhard Schölkopf, Thomas Brox, Peter Gehler。迈向工业异常检测的完全召回。于 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 第14318–14328页, 2022年。2, 3, 5, 6, 8 [23] Marco Rudolph, Bastian Wandt, Bodo Rosenhahn。相似但不同：使用标准化流进行半监督缺陷检测。于 *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, 第1907–1916页, 2021年。3 [24] Marco Rudolph, Tom Wehrbein, Bodo Rosenhahn, Bastian Wandt。用于基于图像的缺陷检测的全卷积跨尺度流。于 *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 第1088–1097页, 2022年。2, 3, 5, 6 [25] Lukas Ruff, Robert Vandermeulen, Nico Goernitz, Lucas Deecke, Shoaib Ahmed Siddiqui, Alexander Binder, Emmanuel Müller, Marius Kloft。深度单类分类。于 *International conference on machine learning*, 第4393–4402页。PMLR, 2018年。8

- [26] Mohammadreza Salehi, Niousha Sadjadi, Soroosh Baselizadeh, Mohammad H Rohban, 与 Hamid R Rabiee。用于异常检测的多分辨率知识蒸馏。收录于 *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 第14902–14912页, 2021年。
- 8[27] Thomas Schlegl, Philipp Seeböck, Sebastian M Waldstein, Ursula Schmidt-Erfurth, 与 Georg Langs。利用生成对抗网络进行无监督异常检测以指导标记发现。收录于 *International conference on information processing in medical imaging*, 第146–157页。Springer出版社, 2017年。
- 8[28] Kihyuk Sohn, Chun-Liang Li, Jinsung Yoon, Minho Jin, 与 Tomas Pfister。深度单类分类的表征学习与评估。收录于 *International Conference on Learning Representations*, 2021年。
- 8[29] Zhou Wang, Alan C Bovik, Hamid R Sheikh, 与 Eero P Simoncelli。图像质量评估：从误差可见性到结构相似性。*IEEE transactions on image processing*, 13(4):600–612, 2004年。
- 2[30] Vitjan Zavrtanik, Matej Kristan, 与 Danijel Skočaj。Draem——一种用于表面异常检测的判别性训练重建嵌入方法。收录于 *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 第8330–8339页, 2021年。
- 2, 4, 5[31] Vitjan Zavrtanik, Matej Kristan, 与 Danijel Skočaj。通过修复重建进行视觉异常检测。*Pattern Recognition*, 112:107706, 2021年。