

# SimpleNet: A Simple Network for Image Anomaly Detection and Localization

## 摘要

本文提出了一个简单且易应用的网络，称为 SimpleNet，来检测和定位异常。SimpleNet 由四个组件组成：（1）一个生成局部特征的预训练特征提取器，（2）一个将局部特征转移到目标域的浅层特征适配器，（3）一个简单的异常特征生成器，通过添加高斯噪声来伪造异常特征正常特征，以及（4）区分异常特征和正常特征的二元异常鉴别器。在推理过程中，异常特征生成器将被丢弃。该方法基于三个直觉。首先，将预先训练的特征转换为面向目标的特征有助于避免领域偏差。其次，在特征空间中生成合成异常更为有效，因为缺陷在图像空间中可能没有太多共性。第三，简单的判别器非常高效且实用。SimpleNet 用简单的方法在数量和质量上都优于以前的方法。在 MVTec AD 基准上，SimpleNet 实现了 99.6% 的异常检测 AUROC，与排名第二的最佳性能模型相比，误差减少了 55.5%。

本报告基于本文提出的方法以及开源代码对 SimpleNet 模型进行复现，对数据预处理方法进行修改后，再次进行相关的模型训练，最终在训练后的结果上分析，实现了与原文接近的模型性能。在原文中异常分割和异常检测的综合结果分别为 99.6% 和 98.1%，在本文中对应 SimpleNet 的实验进行复现中，异常分割和异常检测的综合结果分别为 99.2% 和 98.4%，在异常检测指标上实现与原文相匹配的性能表现，在异常分割指标上则相比原文有所提高。

**关键词：**工业异常检测；异常分割；深度学习

## Abstract

This article proposes a simple and easily applicable network called SimpleNet to detect and locate 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 to the target domain, (3) a simple anomaly feature generator that forges normal features by adding Gaussian noise, and (4) a binary anomaly discriminator that distinguishes between abnormal features and normal features. During the inference process, the anomaly feature generator will be discarded. This method is based on three intuitions. Firstly, converting pre trained features into target oriented features helps to avoid domain bias. Secondly, generating synthetic anomalies in the feature space is more effective, as defects may not have much commonality in the image space. Thirdly, a simple discriminator is highly efficient and practical. SimpleNet outperforms previous methods in both quantity and quality using simple methods. On the MVTec AD benchmark, SimpleNet achieved 99.6% anomaly detection AUROC, with an error reduction of 55.5% compared to the second best performing model.

This report is based on the method proposed in this article and the open-source code to reproduce the SimpleNet model. After modifying the data preprocessing method, relevant model training was conducted again. Finally, the training results were analyzed and the model performance was achieved that was close to the original text. In the original text, the comprehensive results of anomaly segmentation and anomaly detection were 99.6% and 98.1%, respectively. In the reproduction of the experiment on SimpleNet in this article, the comprehensive results of anomaly segmentation and anomaly detection were 99.2% and 98.4%, respectively. The performance of anomaly detection indicators matched the original text, while the performance of anomaly segmentation indicators was improved compared to the original text.

**Keywords:** Industrial anomaly detection; Abnormal segmentation; Deep learning

# 1 引言

图像异常检测和定位任务旨在识别异常图像并定位异常子区域。检测各种感兴趣的异常的技术在工业检测中具有广泛的应用[3,6]。在工业场景中，异常检测和定位尤其困难，因为异常样本稀缺，并且异常可能从细微的变化（例如细划痕）到大的结构缺陷，这种情况使得监督方法并不能有效解决该问题。图1显示了 MVTEC AD 基准[3]的一些示例以及我们提出的方法的结果。

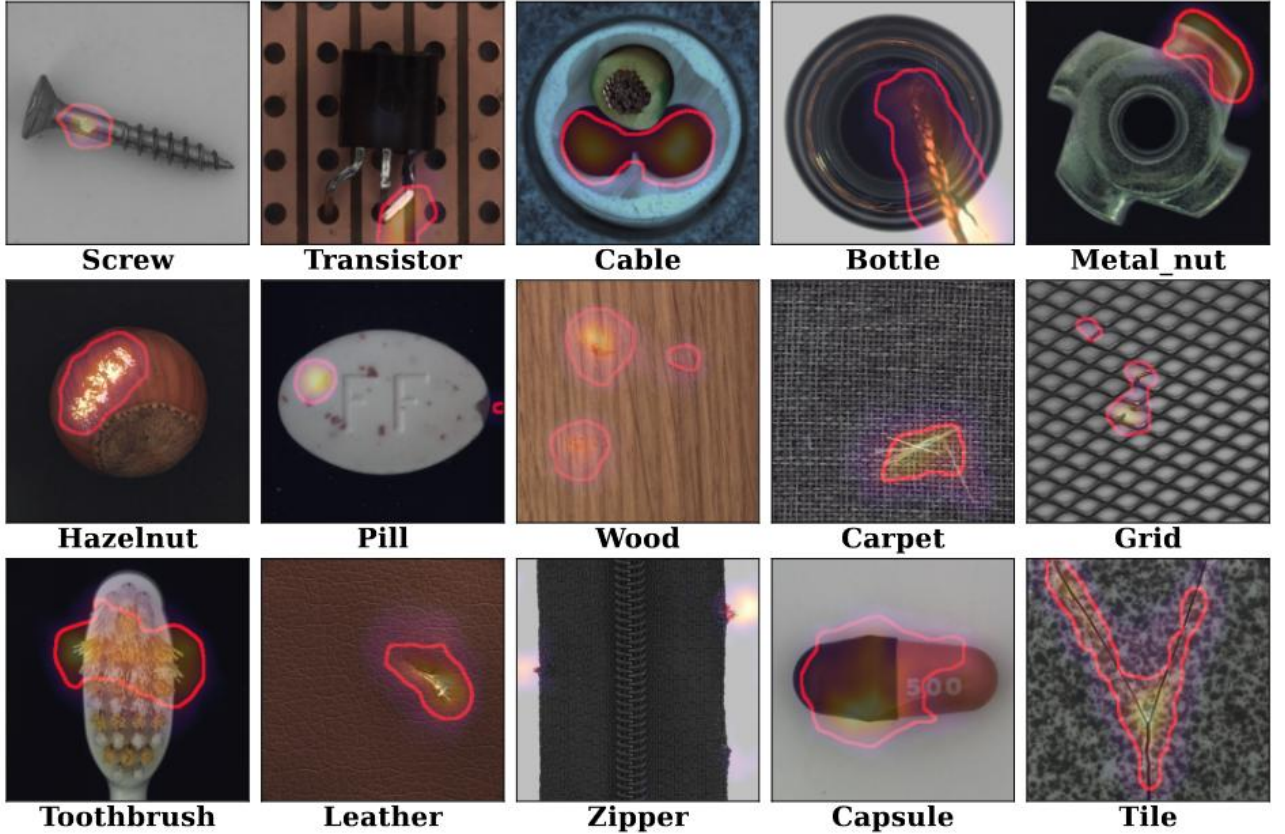


图 1 MVTEC-AD 中样本的可视化。生成的异常图叠加在图像上。高异常分数的异常区域用橙色着色。红色边界表示异常的实际分割图的轮廓。

当前的方法以无监督的方式解决这个问题，在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]和基于嵌入的方法[6,22,24]是解决这个问题的三个主要趋势。基于重建的方法（例如[21,31]）假设仅使用正常数据训练的深度网络无法准确重建异常区域。像素级重建误差被视为异常定位的异常分数。但这种假设可能并不总是成立，有时网络可以很好地“泛化”，以至于它也可以很好地重建异常输入，从而导致误检测[10,19]。基于合成的方法[17,30]通过对无异常图像上生成的合成异常进行训练来估计正常和异常之间的决策边界。然而，合成的图像不够真实。合成数据的特征可能远离正常特征，使用此类负样本进行训练可能会导致松散界限的正常特征空间，这意味着分布内特征空间中可能包含不明确的缺陷。

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) 来提取广义法线特征。然后采用多元高斯分布、归一化流和记忆库等统计算法来嵌入正态特征分布。通过将输入特征与学习的分布或记忆的特征进行比较来检测异常。然而，工业图像通常具有与 ImageNet 不同的分布。直接使用这些有偏差的特征可能会导致不匹配问题。此外，统计算法总是面临高计算复杂度或高内存消耗的问题。

为了缓解上述问题，本文提出了一种新颖的异常检测和定位网络，称为 SimpleNet。SimpleNet利用了基于合成和基于嵌入的方式，并做了一些改进。首先，我们建议使用特征适配器来生成面向目标的特征，从而减少域偏差，而不是直接使用预训练的特征。其次，我们

建议通过向特征空间中的正常特征添加噪声来生成异常特征，而不是通过直接合成图像上的异常。我们认为，通过正确校准噪声尺度，可以获得紧密边界的法线特征空间。第三，我们通过训练一个简单的鉴别器来简化异常检测过程，这比上述基于嵌入的方法采用的复杂统计算法的计算效率要高得多。具体来说，SimpleNet 利用预训练的主干进行正常特征提取，然后使用特征适配器将特征传输到目标域。然后，通过向适应的正常特征添加高斯噪声来简单地生成异常特征。由几层 MLP 组成的简单鉴别器根据这些特征进行训练，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 WideResnet50 主干网，在 MVTec AD 上以 77 fps 运行时实现了 99.6% AUROC，在准确性和效率上超越了之前发布的最佳异常检测方法，见图 2。我们进一步将 SimpleNet 引入到 OneClass 新颖性检测的任务显示其通用性。这些优势使得 SimpleNet 弥合了学术研究和工业应用之间的差距。

## 2 相关工作

异常检测和定位方法主要分为三类，即基于重建的方法、基于合成的方法和基于嵌入的方法。

基于重建的方法认为，异常图像区域不应该能够正确重建，因为它们不存在于训练样本中。一些方法[10]利用生成模型（例如自动编码器和生成对抗网络[11]）来编码和重建正常数据。其他方法[13,21,31]将异常检测视为修复问题，其中图像中的补丁被随机屏蔽。然后，利用神经网络来预测被删除的信息。整合结构相似性指数（SSIM）[29]损失函数在训练中被广泛使用。异常图是作为输入图像与其重建图像之间的像素差异而生成的。然而，如果异常与正常训练数据共享共同的组成模式（例如局部边缘），或者解码器“太强”而无法很好地解码某些异常编码，则图像中的异常可能会得到很好的重建[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对综合生成的刚刚超出分布的模式进行有区别的训练。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，此策略通过剪切图像块并粘贴到大图像的随机位置。CNN 经过训练可以将图像与正常数据分布和增强数据分布区分开来。然而，合成异常的外观与真实异常的外观并不十分吻合。在实践中，由于缺陷多种多样且不可预测，因此不可能生成包含所有异常值的异常集。使用所提出的 SimpleNet，不是合成图像上的异常，而是在特征空间中合成负样本。

基于嵌入的方法最近实现了最先进的性能。这些方法将普通特征嵌入到压缩空间中。异常特征远离嵌入空间中的正常簇。典型方法[6,7,22,24]利用在 ImageNet 上预先训练的网络进行特征提取。PaDiM [6] 通过预先训练的模型，通过多元高斯分布嵌入提取的异常斑块特征。PatchCore[22]使用标称补丁特征的最大代表性内存库。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。直接使用预先训练的特征可能会导致不匹配问题。此外，计算协方差的倒数[6]或搜索内存库中的最近邻居[22]都会限制实时性能，尤其是对于边缘设备。

CS-Flow[24]、CFLOW-AD[12]和DifferNet[23]提出通过归一化流（NF）[20]将正态特征分布转换为高斯分布。由于归一化流只能处理全尺寸的特征图，即不允许下采样，并且耦合层[9]消耗的内存是普通卷积层的几倍，因此这些方法非常消耗内存。蒸馏方法[4,7]训练学生网络，使其与固定的预训练教师网络的输出仅与正常样本相匹配。给出异常查询时，应该检测到学生和教师输出之间的差异。由于输入图像应同时经过教师和学生，因此计算复杂度加倍。

SimpleNet 克服了上述问题。SimpleNet 使用特征适配器在目标数据集上执行迁移学习，以减轻预训练 CNN 的偏差。SimpleNet 建议在特征空间中合成异常，而不是直接在图像上合成。SimpleNet 在推理时遵循单流方式，完全由传统的 CNN 模块构建，有利于快速训练、推理和工业应用。



### 3 本文方法

SimpleNet网络的框架如图2所示。SimpleNet网络首先使用特征适配器来产生面向目标的特征，而不是直接使用预训练的特征来减少域偏差；其次，在特征空间中对正常特征施加噪声来生成异常特征，而不是直接在图像上合成异常特征；第三，通过训练一个简单的鉴别器来简化异常检测过程。具体来说，SimpleNet使用预训练的主干进行常规特征提取，然后使用特征适配器将特征转移到目标域。然后，通过在适应的正态特征上加入高斯噪声，简单地生成异常特征。在这些特征上训练一个由几层MLP组成的简单判别器来判别异常。

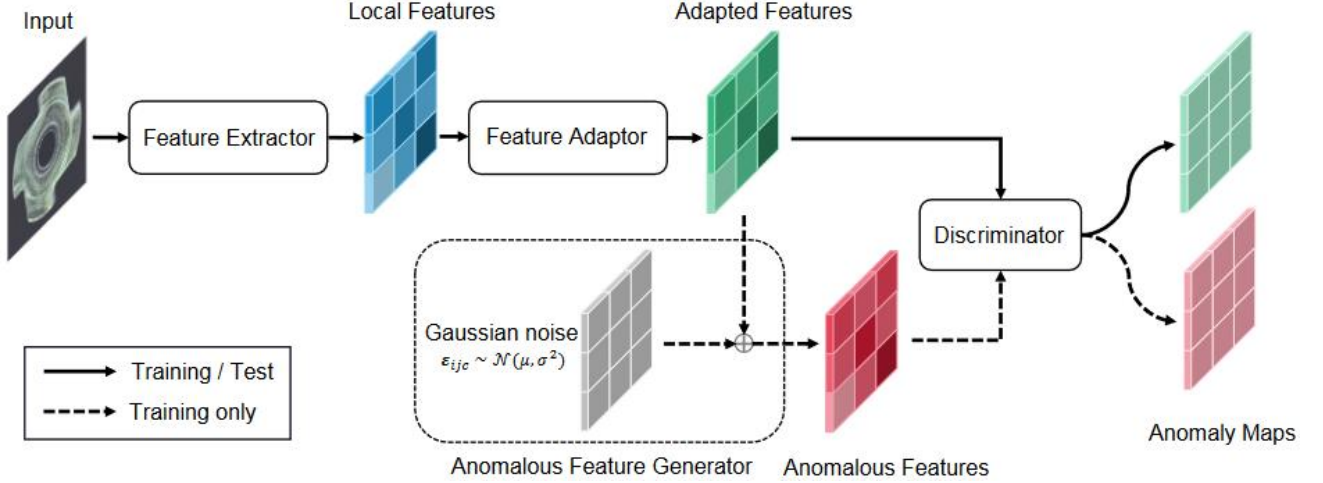


图 2所提出的 SimpleNet 的概述。在训练阶段，标称样本被输入到预先训练的特征提取器中获得局部特征。然后，利用特征适配器将预训练的特征适应目标域。通过向适应特征添加高斯噪声来合成异常特征。适应特征和异常特征分别作为正样本和负样本来训练最终的判别器。异常特征生成器在推理时被删除。

#### 3.1 特征提取器

将训练集和测试集表示为 $X_{train}$ 和 $X_{test}$ 。对于 $X_{train} \cup X_{test}$ 中的任何图像 $x_i \in R^{H \times W \times 3}$ ，预训练网络应从不同层次提取特征，就像通常使用ResNet的主干一样。由于预训练的网络偏向于训练它的数据集，因此只为目标数据集选择一个子集级别是合理的。形式上，我们定义 $L$ 子集，包括层次结构的索引以供使用。层 $l \in L$ 的特征图表示 $\Phi^l, i \sim \Phi^l(x_i) \in R^{H_l \times W_l \times C_l}$ ，其中 $H_l, W_l, C_l$ 分别为特征图的高度、宽度和通道大小。对于位置 $(h, w)$ 处的条目 $\Phi_{h, w}^{l, i} \in R^{C_l}$ ，其具有patchsize  $p$ 的邻域定义为：

$$N_p^{(h, w)} = \left\{ (h', y') \left| \begin{array}{l} h' \in [h - \lfloor \frac{p}{2} \rfloor, \dots, h + \lfloor \frac{p}{2} \rfloor] \\ y' \in [w - \lfloor \frac{p}{2} \rfloor, \dots, w + \lfloor \frac{p}{2} \rfloor] \end{array} \right. \right\} \quad (1)$$

使用聚合函数 $f_{agg}$ 聚合邻域 $N_p^{(h, w)}$ 内的特征得到局部特征 $z_{h, w}^{l, i}$ ，如下：

$$z_{h, w}^{l, i} = f_{agg} \left( \left\{ \Phi_{h', w'}^{l, i} \mid (h', y') \in N_p^{(h, w)} \right\} \right) \quad (2)$$

为了组合来自不同层次的特征 $z_{h, w}^{l, i}$ ，将所有特征图线性调整为相同的大小 $(H_0, W_0)$ ，即最大的特征图的大小。简单地按通道连接特征图，得到特征图 $o^l \in R^{H_0 \times W_0 \times C}$ 。该过程定义为：

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

将上式子简化为

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

其中， $F_\phi$ 是特征提取器。

### 3.2 特征适配器

由于工业图像通常与骨干预训练中使用的数据集分布不同，我们采用特征提取器  $G_\theta$  将训练特征转移到目标域。特征适配器  $G_\theta$  将局部特征  $q_{h,w}$  投影到自适应特征

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

特征适配器可以由简单的神经块组成，如全连接层或多层感知器(MLP)。

### 3.3 异常特征生成器

为了训练鉴别器来估计样本正常的可能性，最简单的方法是对负样本进行采样，并将其与正常样本一起优化。但是，缺陷的缺乏使得采样分布估计变得棘手。SimpleNet通过向正常特征添加高斯噪声来伪造异常特征，在特征空间中合成异常，而不是直接在图像上合成。也就是说训练阶段只使用正常样本，通过高斯噪声产生异常，解决了异常样本不足的问题。

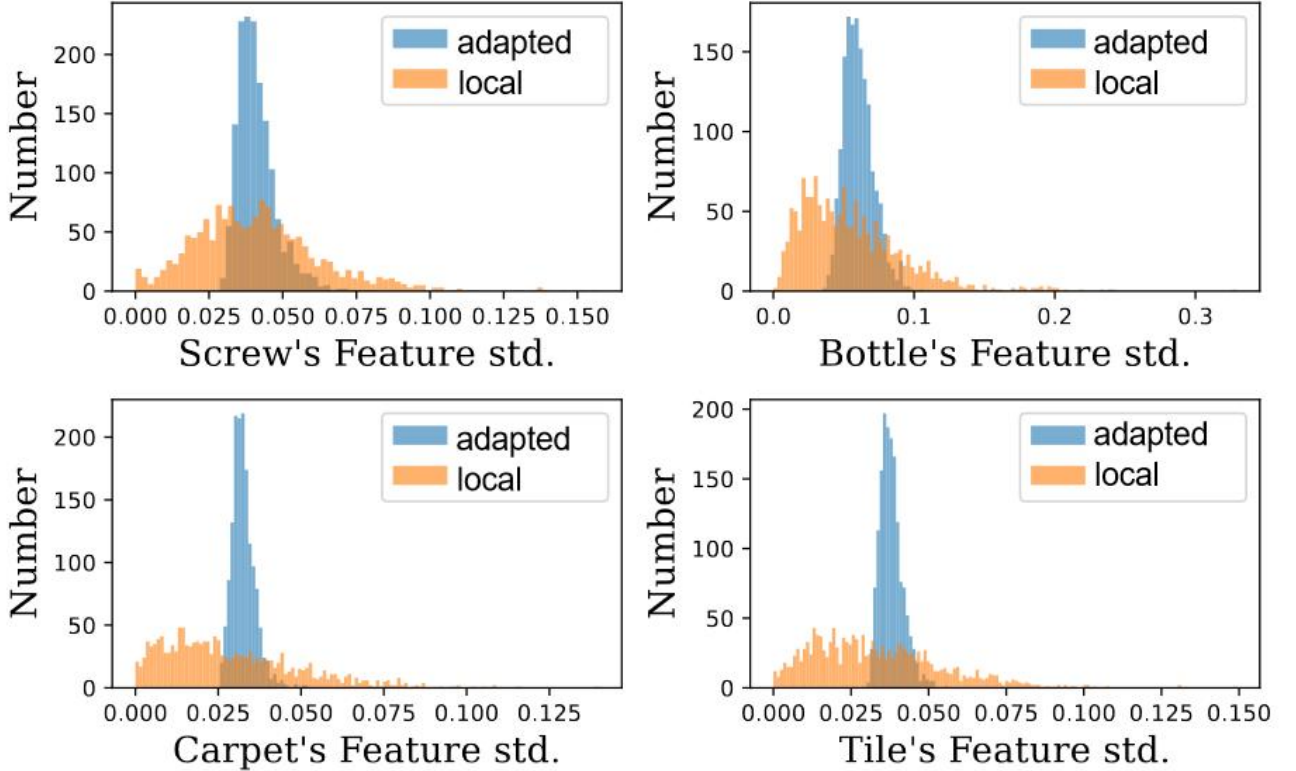


图 3 局部特征和适应特征沿每个维度的标准差直方图。当使用异常特征进行训练时，适应的特征空间变得更加紧凑。

通过在正常特征  $q_{h,w}^i \in \mathbb{R}^C$ ，上添加高斯噪声生成异常特征。形式上，采样一个噪声向量  $\epsilon \in \mathbb{R}^C$ ，每个条目后跟一个 i.i.d. 高斯分布  $N(\mu, \sigma^2)$ 。异常特征  $q_{h,w}^i$  融合为：

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

将经过异常特征训练之后的特征空间与之前进行了比较。可以看到，经过适应的特征在每个维度上的标准差趋于一致。也就是加入伪异常之后的特征空间更加紧凑。

图4说明了异常特征对四类异常监测数据集的影响。我们可以看到，沿着适应特征的每个维度的标准偏差趋于一致。因此，当区分异常特征和正常特征时，特征空间往往是紧凑的。

### 3.4 鉴别器

鉴别器  $D_\psi$  作为正态性记分器，直接估计每个位置  $(h, w)$  的正态性。将正常样本和伪异常样本共同馈送到异常鉴别器中，鉴别器将会正常特征作为正输出，将异常特征作为负输出，这里的鉴别器仅使用一个2层多层感知器（MLP）结构。

### 3.5 损失函数与训练

一个简单的截断损失可推导为：

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

其中,  $th^+$  和  $th^-$  是防止过拟合的截断项，默认选为0.5和-0.5，所以训练的目标就是将正

样本的判别器输出  $D_\psi = q_{h,w}^i$  接近接近0.5，而伪异常的判别器输出  $D_\psi = q_{h,w}^i$  接近-0.5

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

我们将在实验部分用广泛使用的交叉熵损失对提出的截断L1损失函数进行实验评估。训练过程的伪代码如算法1所示。

算法 1 SimpleNet在Pytorch中训练的伪代码

---

```

# 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_)

```

---

### 3.6 异常评分

异常特征生成器在推理时被丢弃。剩余的模块剩余的模块可以堆叠到端到端网络中。我们将每个  $x^i \in X_{\text{test}}$  依次输入上述特征提取器  $G_\theta$  和特征适配器  $G_\theta$  以获得适应的特征  $q_{h,w}^i$ 。异常分数由鉴别器  $D_\psi$  提供为：

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

推理过程中用于异常定位的异常图定义为：

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

然后对  $S_{AL}(x_i)$  进行插值，以获得输入样本的空间分辨率，并对光滑边界进行  $\sigma=4$  的高斯滤波。由于对于任何大小的异常区域都存在最具响应性的点，因此将异常图的最大得分作为每个图像的异常检测得分：

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

## 4 复现细节

### 4.1 数据集

本文的实验是在 MVTEC 异常检测基准 [3] 上进行的，即异常检测和定位领域著名的数据集。MVTEC AD 包含源自制造的 5 个纹理和 10 个对象类别，总共 5354 个图像。该数据集由用于训练的正常图像和用于测试的具有各种类型缺陷的正常图像和异常图像组成。它还为有缺陷的测试图像提供像素级注释。典型图像如图 1 所示。

对于图像的预处理，原文中分别对图像调整大小和中心裁剪为  $256 \times 256$  和  $224 \times 224$ ，并且不应用数据增强。本文中观察到纹理类别的图像覆盖整个图像，故为了使输入数据具有更为完整的原始信息，对于纹理类别图像直接调整大小为  $256 \times 256$ 。同时，本文遵循一类分类协议，也称为冷启动异常检测，本文在其各自的正常训练样本上为每个类别训练一类分类器。

### 4.2 评估指标

图像级异常检测性能通过标准接收器算子曲线下面积（我们将其表示为 I-AUROC）进行测量，并使用生成的异常检测分数 SAD（公式 11）。对于异常定位，异常图（方程 10）用于评估像素级 AUROC（表示为 P-AUROC）。根据先前的工作[6,22]，我们在 MVTEC AD 上计算用于检测和定位的类平均 AUROC 和平均 AUROC 总体类别。在原文中，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]。

### 4.3 实施细节

本文实验基于原文在网上的公开代码进行修改。本文实验中使用的所有主干网络均使用 ImageNet[8]进行了预训练。主干的第二和第三中间层，例如当主干网是 ResNet 的架构时，等式 3 中的  $l' \in [2, 3]$  用于特征提取器，如 [22] 中所示。默认情况下，我们的实现使用 WideResnet50 作为主干，特征提取器的特征维度设置为 1536。后面的特征适配器本质上是一个没有偏差的全连接层。适配器中 FC 层的输入和输出特征的尺寸相同。异常特征生成器添加了 i.i.d. 高斯噪声  $N(0, \sigma^2)$  到每个正常特征条目。 $\sigma$  默认设置为 0.015。随后的判别器由线性层、批量归一化层、leaky relu（0.2 斜率）和线性层组成。在等式 7 中  $th+$  和  $th-$  均设置为 0.5。使用 Adam 优化器，将特征适配器和鉴别器的学习率分别设置为 0.0001 和 0.0002，并将权重衰减设置为 0.00001。每个数据集的训练 epochs 设置为 160，batchsize 为 4。



## 5 实验结果分析

MVTec AD 上的异常检测结果如表 1 所示。图像级异常得分由异常图的最高得分给出，如公式 11 所示。SimpleNet 在 15 个类别中的 9 个类别中获得了最高得分。在原文中，对于纹理和对象，SimpleNet 分别达到了 I-AUROC 的 99.8% 和 99.5% 的新 SOTA（在表1中为Type为Original的列所示结果）。在本文中，基于原作者给出的代码进行修改复现后，对于纹理和对象，SimpleNet 分别达到了 I-AUROC 的 99.9% 和 98.8% 的结果。原文和本文的SimpleNet 实现了显着更高的平均图像异常检测性能，即 I-AUROC 得分为 99.6%。请注意，PatchCore[22]（下一个最佳竞争对手，在相同的 WideResnet50 主干网下）的误差从 0.9% 减少到 SimpleNet 的 0.4%，意味着误差减少了 55.5%。在工业检测环境中，这是一个相关且显著的减少。

表 1 SimpleNet原文结果和复现结果与 MVTec-AD上最先进的工作的比较。图像级AUROC (I-AUROC) 和像素级 AUROC (P-AUROC) 在每个条目中显示为 I-AUROC%/P-AUROC%。CS-Flow 的 P-AUROC 未记录在[24]中

Type	Reconstruction-based		Synthesizing-based		Embedding-based				Original	Ours
Model	AE-SSIM	RIAD	DRaEM	CutPaste	CS-Flow	PaDiM	RevDist	PatchCore	SimpleNet	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	99.4/98.9
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	100/98.8
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	100/96.5
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	99.9/94.2
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	100/99.2
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	99.9/97.5
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	100/98.6
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	99.9/98.6
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	99.7/99.3
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	100/98.7
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	100/99.1
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	98.9/99.0
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	96.1/99.4
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	93.6/98.7
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	100/98.4
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	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	98.8/98.9
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	<b>99.6/98.1</b>	<b>99.2/98.4</b>

异常定位性能是通过像素级 AUROC 来衡量的，记为 P-AUROC。与最先进方法的比较如表 1 所示。在原文中，SimpleNet 在 MVTec AD 上实现了98.1%的P-AUROC 的最佳异常检测性能，并且在对象类别上实现了 98.4% P-AUROC 的新 SOTA。在本文中，针对SimpleNet 的复现，使SimpleNet在 MVTec AD 上实现了98.4% 的 P-AUROC 的最佳异常检测性能，实现了新 SOTA，并且在对象类别上实现了 98.9% P-AUROC 的新 SOTA，同时在纹理类别上实现了与原文中97.5%相匹配的性能表现。

在原文中，SimpleNet 在 15 个类别中的4个类别中取得了最高分。而在本文中，SimpleNet 在 15 个类别中的5个类别中取得了最高分。为了更为直观地观察该方法实现的异常检测效果，我们将异常定位的代表性样本可视化，如图4和图5所示。



## 6 总结与展望

原文中介绍了一种名为SimpleNet的简单而有效的方法，用于在无监督情况下进行异常检测和定位。SimpleNet由多个简单的神经网络模块组成，这使得它易于训练并可以在工业场景中得以应用。尽管其设计相对简单，但在MVtec AD基准测试中，SimpleNet相较于先前最先进的方法表现出了最优异的性能。这为异常检测和定位领域提供了一个新的视角，有助于弥合学术研究和工业应用之间的差距。

为了验证原文的有效性，本报告基于本文提出的方法以及开源代码对SimpleNet模型进行复现，对数据预处理方法进行修改后，再次进行相关的模型训练，最终在训练后的结果上分析，实现了与原文接近的模型性能。在原文中异常分割和异常检测的综合结果分别为99.6%和98.1%，在本文中对SimpleNet的实验进行复现中，异常分割和异常检测的综合结果分别为99.2%和98.4%，在异常检测指标上实现与原文相匹配的性能表现，在异常分割指标上则相比原文有所提高。这表明SimpleNet不仅在研究领域表现出色，而且在实际应用中也具有竞争力。这一成果有望推动无监督异常检测和定位方法在工业环境中的广泛应用，为相关领域的发展提供了有力支持。

## 参考文献

- [1] Davide Abati, Angelo Porrello, Simone Calderara, and Rita Cucchiara. Latent space autoregression for novelty detection. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pages 481–490, 2019. 8
- [2] Jinwon An and Sungzoon Cho. Variational autoencoder based anomaly detection using reconstruction probability. Special Lecture on IE, 2(1):1–18, 2015. 8
- [3] Paul Bergmann, Michael Fauser, David Sattlegger, and Carsten Steger. Mvtec ad—a comprehensive real-world dataset for unsupervised anomaly detection. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pages 9592–9600, 2019. 1, 5
- [4] Paul Bergmann, Michael Fauser, David Sattlegger, and Carsten Steger. Uninformed students: Student-teacher anomaly detection with discriminative latent embeddings. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 4183–4192, 2020. 3
- [5] Yuanhong Chen, Yu Tian, Guansong Pang, and Gustavo Carneiro. Deep one-class classification via interpolated gaussian descriptor. In Proceedings of the AAAI Conference on Artificial Intelligence, pages 383–392, 2022. 8
- [6] Thomas Defard, Aleksandr Setkov, Angelique Loesch, and Romaric Audigier. Padim: a patch distribution modeling framework for anomaly detection and localization. In International Conference on Pattern Recognition, pages 475–489. Springer, 2021. 1, 2, 3, 5, 8
- [7] Hanqiu Deng and Xingyu Li. Anomaly detection via reverse distillation from one-class embedding. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022. 2, 3, 5, 8
- [8] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In 2009 IEEE conference on computer vision and pattern recognition, pages 248–255. Ieee, 2009. 5
- [9] Laurent Dinh, Jascha Sohl-Dickstein, and Samy Bengio. Density estimation using real nvp. arXiv preprint arXiv:1605.08803, 2016. 3
- [10] Dong Gong, Lingqiao Liu, Vuong Le, Budhaditya Saha, Moussa Reda Mansour, Svetha Venkatesh, and Anton van den Hengel. Memorizing normality to detect anomaly: Memory-augmented deep autoencoder for unsupervised anomaly detection. In Proceedings of the IEEE/CVF International Conference on Computer Vision, pages 1705–1714, 2019. 2
- [11] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. Advances in neural information processing systems, 27, 2014. 2
- [12] Denis Gudovskiy, Shun Ishizaka, and Kazuki Kozuka. Cflow-ad: Real-time unsupervised anomaly detection with localization via conditional normalizing flows. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pages 98–107, 2022. 3
- [13] Matthias Haselmann, Dieter P Gruber, and ul Tabatabai. Anomaly detection using deep learning based image completion. In 2018 17th IEEE international conference on machine learning and applications (ICMLA), pages 1237–1242. IEEE, 2018. 2
- [14] Jinlei Hou, Yingying Zhang, Qiaoyong Zhong, Di Xie, Shiliang Pu, and Hong Zhou. Divide-and-assemble: Learning block-wise memory for unsupervised anomaly detection. In Proceedings of the IEEE/CVF International Conference on Computer Vision, pages 8791–8800, 2021. 8
- [15] Wenpeng Hu, Mengyu Wang, Qi Qin, Jinwen Ma, and Bing Liu. Hrn: A holistic approach to one class learning. Advances in Neural Information Processing Systems, 33:19111–19124, 2020. 8
- [16] A Krizhevsky. Learning multiple layers of features from tiny images. Master’s thesis, University of Tront, 2009. 5, 8
- [17] Chun-Liang Li, Kihyuk Sohn, Jinsung Yoon, and Tomas Pfister. Cutpaste: Self-supervised learning for anomaly detection and localization. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 9664–9674, 2021. 2, 4, 5
- [18] Philipp Liznerski, Lukas Ruff, Robert A Vandermeulen, Billy Joe Franks, Marius Kloft, and Klaus Robert Muller. Explainable deep one-class classification. In International Conference on Learning Representations, 2020. 4
- [19] Pramuditha Perera, Ramesh Nallapati, and Bing Xiang. Ogan: One-class novelty detection using gans with constrained latent representations. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 2898–2906, 2019. 2, 8
- [20] Danilo Rezende and Shakir Mohamed. Variational inference with normalizing flows. In International conference on machine learning, pages 1530–1538. PMLR, 2015. 3
- [21] Nicolae-C at alin Ristea, Neelu Madan, Radu Tudor Ionescu, Kamal Nasrollahi, Fahad Shahbaz Khan, Thomas B Moeslund, and Mubarak Shah. Self-supervised predictive convolutional attentive block for anomaly detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 13576–13586, 2022. 2
- [22] Karsten Roth, Latha Pemula, Joaquin Zepeda, Bernhard Scholkopf, Thomas Brox, and Peter Gehler. Towards total recall in industrial anomaly detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 14318–14328, 2022. 2, 3, 5, 6, 8
- [23] Marco Rudolph, Bastian Wandt, and Bodo Rosenhahn. Same same but different: Semi-supervised defect detection with normalizing flows. In Proceedings of the IEEE/CVF winter conference on applications of computer vision, pages 19071916, 2021. 3



- [24] Marco Rudolph, Tom Wehrbein, Bodo Rosenhahn, and Bastian Wandt. Fully convolutional cross-scale-flows for imagebased defect detection. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 1088–1097, 2022. 2, 3, 5, 6
- [25] Lukas Ruff, Robert Vandermeulen, Nico Goernitz, Lucas Deecke, Shoaib Ahmed Siddiqui, Alexander Binder, Emmanuel Muller, and Marius Kloft. Deep one-class classification. In *International conference on machine learning*, pages 4393–4402. PMLR, 2018.
- [26] Mohammadreza Salehi, Niousha Sadjadi, Soroosh Baselizadeh, Mohammad H Rohban, and Hamid R Rabiee. Multiresolution knowledge distillation for anomaly detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 14902–14912, 2021. 8
- [27] Thomas Schlegl, Philipp Seeböck, Sebastian M Waldstein, Ursula Schmidt-Erfurth, and Georg Langs. Unsupervised anomaly detection with generative adversarial networks to guide marker discovery. In *International conference on information processing in medical imaging*, pages 146–157. Springer, 2017. 8
- [28] Kihyuk Sohn, Chun-Liang Li, Jinsung Yoon, Minho Jin, and Tomas Pfister. Learning and evaluating representations for deep one-class classification. In *International Conference on Learning Representations*, 2021. 8
- [29] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004. 2
- [30] Zavrtnik V, Kristan M, Skočaj D. Draem-a discriminatively trained reconstruction embedding for surface anomaly detection[C]//*Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2021: 8330-8339.
- [31] Zavrtnik V, Kristan M, Skočaj D. Reconstruction by inpainting for visual anomaly detection[J]. *Pattern Recognition*, 2021, 112: 107706.