

利用多层次特征融合网络的图像异常检测算法

唐俊¹, 左金梅¹, 王科², 张艳¹, 王年¹

(1. 安徽大学 电子信息工程学院, 安徽 合肥 230031; 2. 安徽大学 互联网学院, 安徽 合肥 230031)

摘要: 图像异常检测旨在识别并定位图像中的异常区域, 针对现有算法中不同层次特征信息利用不充分的问题, 提出了基于多层次特征融合网络的图像异常检测算法。通过使用融合了异常先验知识的伪异常数据生成算法, 对训练集进行了异常数据扩充, 将异常检测任务转化为监督学习任务; 构建了多层次特征融合网络, 将神经网络中不同层次特征进行融合, 丰富了特征中的低层纹理信息和高层语义信息, 使得用于异常检测的特征更具区分性; 训练时, 设计了分数约束损失和一致性约束损失, 并结合特征约束损失对整个网络模型进行训练。实验结果表明, MVTEC 数据集上图像级检测平均 AUROC 值为 98.7%, 像素级定位平均 AUROC 值为 97.9%, 平均 PRO 值为 94.2%, 均高于现有的异常检测算法。

关键词: 图像异常检测; 伪异常; 多层次特征融合; 一致性约束

中图分类号: TN911.73; TP301.6 文献标识码: A

Image anomaly detection algorithm using multi-level feature fusion network

TANG Jun¹, ZUO Jinmei¹, WANG Ke², ZHANG Yan¹, WANG Nian¹

(1. School of Electronic Information Engineering, Anhui University, Hefei 230031, China; 2. School of Internet, Anhui University, Hefei 230031, China)

Abstract: Image anomaly detection aims to identify and locate the abnormal region in an image. To address the issue on the insufficient utilization of different-level feature information in the existing methods, an image anomaly detection method based on multi-level feature fusion network is put forward. By using the pseudo anomaly data generation algorithm incorporated with the anomaly prior knowledge, the anomaly data of the training set are augmented, and then the anomaly detection task is transformed into a supervised learning task. A multi-level feature fusion network is constructed to enriches the low-level texture information and high-level semantic information of features by fusing the different levels of features in the neural network, which can make the features used for anomaly detection more discriminative. In the training phase, the score constraint loss and the consistency constraint loss are designed and combined with the feature constraint loss to train the whole network model. Experimental results on the MVTEC dataset show that the proposed model can achieve 98.7% AUROC in the detection task, 97.9% AUROC and 94.2% PRO in the localization task, which outperforms several existing anomaly detection approaches.

Keywords: Image anomaly detection; Pseudo-anomaly; Multi-level feature fusion; Consistency constraint

异常检测通常是指在异常先验知识有限甚至缺失的情况下, 检测并定位出与正常模式不同的异常模式。在计算机视觉领域中, 图像中的异常通常表现出局部性, 如物体表面的划痕、破损等或者因部分结构扭曲而显示出的缺陷, 很多应用任务都涉及异常检测问题, 如工业图像缺陷检测^[1]、医学图像分析^[2,3]和视频监控^[4-6]等。与常见的监督分类问题^[7,8]不同, 图像中的异常模式相比于正常模式十分罕见, 在形式上呈现多样性, 因此难以获取

足够的有标签或无标签的异常数据^[9]用于模型训练。为了解决这一挑战, 研究者们通常仅使用正常数据来训练异常检测器, 现有的检测算法总体上可分为三类: 基于重构的方法^[10-13]、基于概率预测的方法^[14-18]以及基于单类分类^[19-23]的方法。

基于重构和基于概率预测的方法, 在训练时模型通常仅学习正常示例的特征表示或数据分布, 测试时那些无法被模型很好表示的示例将被视为异常。然而, Hendrycks 等人^[23]表明, 除了正常数

基金项目: 国家自然科学基金(61772032); 安徽省重点研究与开发计划项目(202004a07020050); 赤足足迹检索关键技术研究与应用, 安徽省重点研究与开发计划项目-科技合作专项-科技强警(2022k07020006); 基于足底压力的身份识别技术与方法研究, 安徽省高校自然科学研究重大项目(KJ2021ZD0004)

作者简介: 唐俊 (1977—), 男, 安徽黄山人, 教授, 博士, 博士生导师, E-mail: tangjunahu@163.com;

王科 (通信作者), 男, 安徽滁州人, 讲师, 博士, E-mail: wangke4747@126.com

据外，在训练时引入一些已知的异常来增加异常先验知识，可以显著地提高异常检测性能。基于单类分类的方法^[19]，旨在学习一个具有判别力的特征空间，将仅包含正常模式的特征区域与所有的异常特征区域分离开来。Ruff 等人^[21]提出完全卷积描述算法(Fully Convolutional Data Description, FCDD)，该算法在训练时使用正常数据和伪异常数据，通过交叉熵分类损失训练一个超球体，实现将正常数据特征聚集在超球体球心附近，异常数据特征映射到远离球心的区域，测试阶段利用数据点到球心的距离来生成相应的异常分数。然而，该算法存在着如下三个缺点：(1)所使用的伪异常数据类型单一，使得模型在测试阶段无法很好地处理复杂多样的真实异常模式；(2)所使用特征的质量不高，缺乏对图像底层纹理信息的利用；(3)直接通过将数据点到球心的距离转化为异常分数，使得模型性能严重依赖于数据的特征质量，当特征质量不高时，模型性能十分有限。

为了解决 FCDD 算法存在的上述缺点，本文构建了一个多层次特征融合网络，用于提取和融合不同类型的特征，该网络主要包括以下三个部分：主干网络、统计纹理特征提取网络^[24](Statistical Texture Learning Network, STLNet)以及自校准卷积模块^[25](Self-calibrated Convolutions Block)。另外，与 FCDD 不同，本文通过将数据特征经过一个全连接层映射的方式生成相应的异常分数，使得模型能够很好地应对特征质量不高的情况。本文方法具体步骤如下：首先通过自监督方式在正常训练图像上生成伪异常，同步训练多层次特征融合网络和一个单类分类器，使正常数据特征集中在设定的超球球心附近，生成的伪异常数据特征映射在尽可能远离球心的区域；然后使用一个全连接层将正常数据和异常数据特征映射为异常分数，并使得正常数据的异常得分小，异常数据的异常得分大；最后考虑到数据特征与分数之间应保持相类似的拓扑结构，提出了特征与分数之间一致性约束项，用于最终损失函数的设计。

1 多层次特征融合网络

本文构建了一个多层次特征融合网络，如图 1 所示，该网络整体上可以分为主干网络、纹理特征提取网络(STLNet)^[24]和自校准模块^[25]三个部分。通过将 STLNet 提取的底层纹理特征和自校准模块生成的高层语义特征进行有效地融合，得到本文模型最终的输出特征。

其中主干网络为 ResNet^[26]，一个多尺度特征提取器，它的低层网络生成具有较少上下文信息的高分辨率特征，高层网络生成具有较多上下文信息的低分辨率特征。对于 STLNet，它由**纹理增强模块(Texture Enhancement Module, TEM)**和**金字塔纹理特征提取模块(Pyramid Texture Feature Extraction Module, PTFEM)**两个模块组成，该网络首先融合 ResNet 前两层网络输出的特征，并送入到 TEM 模块中提取一级纹理特征；然后将 TEM 模块提取的纹理特征和 ResNet 前两层网络输出的特征进行融合，输入到 PTFEM 模块抽取多尺度纹理特征；最后融合 ResNet 前两层网络输出的特征、一级纹理特征和多尺度纹理特征获取最终的图像底层纹理特征。自校准模块通过下采样-上采样操作间接地增大了卷积核的感受野，本文通过将 ResNet 第三层网络输出的特征输入到自校准模块中获得更高质量的特征。

1.1 STLNet 纹理特征提取网络

STLNet 从统计学角度学习图像纹理特征，提出了使用量化和计数算子来描述和统计图像中纹理信息的方法，分析了图像底层信息的分布。考虑到在异常检测任务中，图像中不同纹理之间的差异性，本文通过 STLNet 提取底层纹理特征，有效地刻画了图像中正常和异常区域的纹理，提高了异常检测的性能。STLNet 设计了 TEM 和 PTFEM 两个模块，分别使用一维量化计数算子(Quantization and Counting Operator, QCO)和二维量化计数算子描述图像中的纹理分布。对于 TEM，如图 2 所示，它逐个考量每个像素在特征空间中的分布情况，首先通过计算输入特征图中的每个像素特征与其全局特征的余弦距离，得到特征相似度向量 $\mathbf{M} \in \mathbb{R}^{1 \times K}$ ，其中 $K = H \times W$ ， H 和 W 分别表示输入特征图的高度和宽度；然后使用一维 QCO 将 $\mathbf{M}$ 量化成 N 个级别，得到量化矩阵 $\mathbf{E} \in \mathbb{R}^{N \times K}$ 和量化级别 $\mathbf{L} = [\mathbf{L}_1, \mathbf{L}_2, \dots, \mathbf{L}_N]$ ，其中

$$\mathbf{L}_n = (\mathbf{M}_{\max} - \mathbf{M}_{\min}) \div N \times n + \mathbf{M}_{\min}, \quad \mathbf{E}_i \in \mathbb{R}^N (i \in [1, K])$$

表示向量 $\mathbf{M}$ 第 i 像素相似值 $\mathbf{M}_i$ 所属的量级级别， $\mathbf{E}_{i,n}$ 的计算公式如下：

$$\mathbf{E}_{i,n} = \begin{cases} 1 - |\mathbf{L}_n - \mathbf{M}_i| & \text{if } -\frac{0.5}{N} \leq \mathbf{L}_n - \mathbf{M}_i < \frac{0.5}{N} \\ 0 & \text{else} \end{cases} \quad (1)$$

将量化级 $\mathbf{L}$ 和归一化后的量化矩阵 $\mathbf{E}'$ 融合生成量

化映射矩阵 $\mathbf{B} \in \mathbb{R}^{N \times 2}$ 。 $\mathbf{B}$ 的计算公式为：

$$\mathbf{B} = \text{Cat} \left(\mathbf{L}, \frac{\sum_{i=1}^K \mathbf{E}_{i,n}}{\sum_{n=1}^N \sum_{i=1}^K \mathbf{E}_{i,n}} \right) \quad (2)$$

最后将量化映射 $\mathbf{B}$ 通过多层感知机 (Multilayer Perceptron, MLP) 增加通道数并与全局特征相融合获得统计特征 $\mathbf{D} \in \mathbb{R}^{N \times C}$ ，其中 C 表示通道数。由于直接从 ResNet 网络浅层提取的底层特征质量通常较低，因此 TEM 利用自注意力机制设计了一个可学习的映射矩阵，该矩阵融合所有其他节点特征来更新统计特征 $\mathbf{D}$ ，得到重构的 $\mathbf{D}'$ ；

然后将特征 $\mathbf{D}'$ 的信息平均分配给输入的所有像素，生成高质量的纹理特征 $\mathbf{R} = \mathbf{D}' \cdot \mathbf{E}$ 。

对于 PTFEM，如图 3 所示，它使用二维 QCO 来逐对考量两个相邻像素在特征空间中的分布情况，旨在计算特征图相邻像素之间的共生关系，是对一维 QCO 的扩展。此外，PTFEM 还使用四个不同尺度 [1, 2, 4, 8] 的并行分支来提取多尺度纹理特征。每个分支上的特征图被划分为不同数量的子区域，每个子区域通过纹理特征提取单元得到相对应的纹理表示；然后将每个分支所得的特征图上采样至输入特征图大小，并进行特征融合，生成多尺度纹理特征。

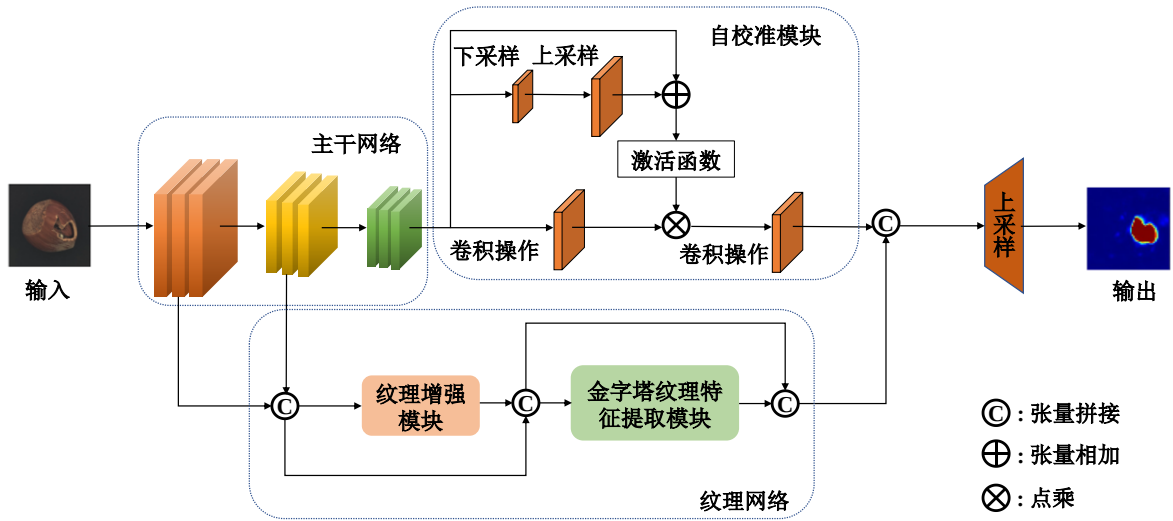


图 1 多层次特征融合网络图

Fig.1 Multi-level feature fusion network diagram

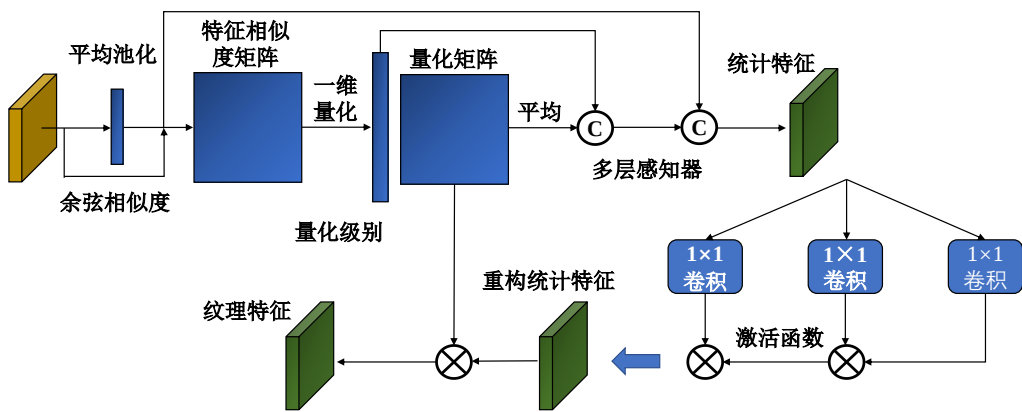


图 2 纹理特征增强模块

Fig2 Texture enhancement module

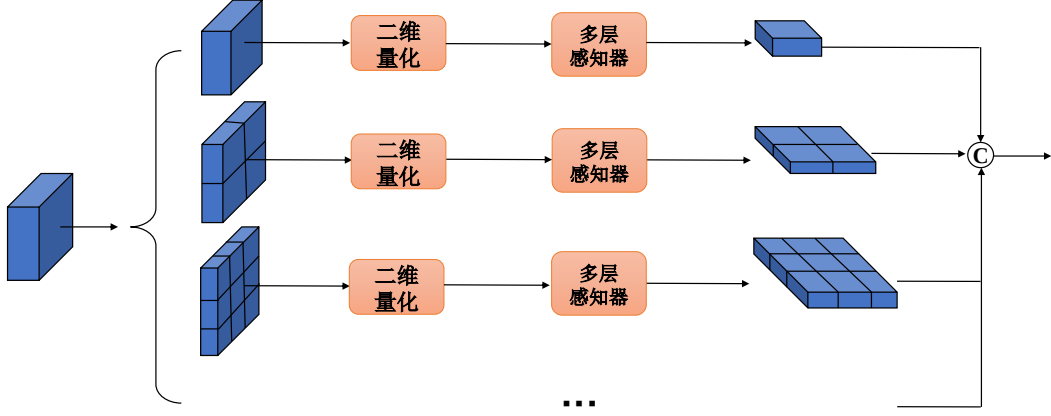


图3 金字塔特征提取模块

Fig.3 Pyramid texture feature extraction module

1.2 自校准模块

本文的主干网络只使用了 ResNet 的前三层网络，导致主干网络的感受野较小。为了获得更大感受野网络下的数据特征，且不改变特征图的高度和宽度，本文使用了一个自校准卷积模块。将 ResNet 第三层网络输出的特征送入自校准卷积模块，通过对特征图进行下采样间接增加卷积核的尺寸，以此扩大网络的感受野，进而生成语义信息更为丰富的特征。

自校准模块在两个不同的尺度空间进行卷积变换，高效地收集了每个空间位置的上下文信息。如图 1 所示，在原始尺度空间中，输入特征经过 Conv-BN 操作(Conv 表示卷积操作，BN 表示批次归一化操作)，不改变特征图宽度和高度；在小尺度特征空间中，输入特征首先经过平均池化下采样(下采样率为 r)后通过 Conv-BN 操作，然后上采样到原始特征图尺寸，并与原始特征进行相加融合，经过 Sigmoid 激活后与原始尺度空间中特征相乘并通过 Conv-BN 操作获得最终的输出。

2 伪异常与损失

2.1 伪异常生成

在异常检测任务中，难以获取充足的异常数据，因此研究人员通常仅仅使用正常数据来训练模型。然而 Ruff 等人^[21]的研究表明，在训练阶段引入一些人为生成的异常数据可以有效地提升模型的检测性能。受此启发，本文采用了两种伪异常生成算法在正常训练图像上生成伪异常，并为每张伪异常图像生成相应的掩膜^[21](Mask)，如图 4 所示。其一为使用“五彩纸屑噪声”生成合成异常，其二为通过 CutPaste^[17]，随机剪切图像上一小块

区域然后贴在图像上的随机位置制造异常。虽然生成的伪异常与真实异常存在着不同，但是本文实验表明，在训练集中引入伪异常数据，可以促使模型对不规则性的学习，使得模型在处理真实缺陷数据时具有很好的泛化能力。

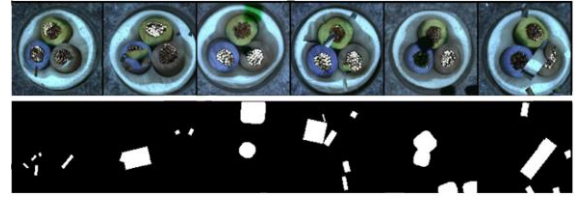


图4 伪异常生成图

Fig.4 Pseudo exception generating graph

2.2 特征约束

本文通过改进超球分类损失(Hypersphere Classifier, HSC)^[21,22]提出了正常-异常特征约束损失，如图 5 所示，该约束项旨在学习一个超球体，通过对神经网络 $\psi(\cdot, \mathbf{w})$ 的训练，将正常数据特征

聚集在球心附近，异常数据特征映射远离球心^[19]。因为图像中的异常通常是细粒度的，为了更准确地分割出图像中的缺陷区域，这里采用密集的像素级预测，即预测出每个像素所属的类别(正常或异常)。设 $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N$ 表示图像像素样本集合，

$y_1, y_2, \dots, y_N$ 表示每个像素所对应的标签，其中

$y_i = 0$ 表示正常、 $y_i = 1$ 表示异常。特征约束损失

表示为：

$$\mathcal{L}_t = \frac{1}{N} \sum_{i=1}^N (1 - y_i) A(\mathbf{x}_i) - y_i \ln(1 - \exp(-A(\mathbf{x}_i))) \quad (3)$$

其中 $A(\mathbf{x}_i) = \sqrt{\psi(\mathbf{x}_i, \mathbf{W})^2 + 1} - 1$ ，表示像素 $\mathbf{x}_i$ 到超球体球心的距离，本文使用原点作为球心， ψ 表示特征学习网络， $\psi(\mathbf{x}_i, \mathbf{W}) \in \mathbb{R}^Q$ 表示像素 $\mathbf{x}_i$ 的特征向量， $\mathbf{W}$ 为网络中的参数。通过最小化 $\mathcal{L}_t$ 学习一个超球体，使得正常像素特征集中在超球体内，异常像素特征映射到超球体外。通过 $l(z) = \exp(-A(\mathbf{x}_i))$

将像素 $\mathbf{x}_i$ 到球心的距离 $A(\mathbf{x}_i)$ 转换为像素属于正常类别的概率，然后使用交叉熵损失训练网络使得正常像素到球心的距离尽可能小，异常像素到球心的距离尽可能大。

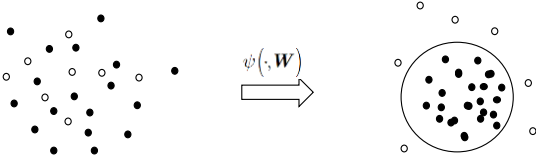


图 5 超球体分类示意图

Fig.5 Diagram of classification of hypersphere

在异常检测与分割任务中，训练集往往不包含异常数据，本文通过引入伪异常生成技术，一定程度上解决了该问题。但是，本文所生成的异常图像中的异常像素数量远远小于正常像素数量，这带来了类别不平衡问题。为了缓解类别不平衡对模型训练阶段的影响，本文为正常和异常特征损失引入不同的权重，重新设计了如下特征约束损失：

$$\mathcal{L}_t = \frac{1}{2N_1} \sum_{i=1}^{N_1} (1 - y_i) A(\mathbf{x}_i) + \frac{1}{2N_2} \sum_{j=1}^{N_2} -y_j \ln(1 - \exp(-A(\mathbf{x}_j))) \quad (4)$$

其中 N_1 表示训练图像中正常像素个数， N_2 表示异常像素个数。

2.3 分数约束

如 2.2 节所述，本文所提出的方法将正常像素

的特征统一映射到超球体内，异常像素的特征映射到超球体外，在特征层面实现对正常像素和异常像素进行区分。但在图像异常分割任务中，还需要预测每个像素的异常得分用于判断其所属的类别。与 FCDD^[21]直接将像素特征点到超球体球心的距离转化成异常分数的方法不同，本文使用一个全连接层 $\eta(\cdot, \theta_s): \mathbf{F} \in \mathbb{R}^Q \mapsto S \in \mathbb{R}$ ，将像素特征 $\mathbf{F}$

映射为异常得分 S ，其中 θ_s 为网络中的参数，有效地应对所得数据特征质量不高的问题。然后设计分数约束项使得正常像素异常得分低，异常像素异常得分高。分数约束损失表示为：

$$\mathcal{L}_s = \frac{1}{N_n} \sum_{i=1}^{N_n} \max(0, S(\mathbf{x}_{i_n}) - \gamma) + \frac{1}{N_a} \sum_{j=1}^{N_a} \max(0, \lambda - (S(\mathbf{x}_{j_a}) - S(\mathbf{x}_{i_n}))) \quad (5)$$

其中 N_n 和 N_a 分别表示正常像素个数和异常像素个数， $\mathbf{x}_{i_n}$ 和 $\mathbf{x}_{j_a}$ 分别表示正常像素和异常像素，

$S(\mathbf{x}_i) = \eta(\psi(\mathbf{x}_i, \mathbf{W}), \theta_s)$ 表示像素 $\mathbf{x}_i$ 的异常得分，

$\psi(\mathbf{x}_i, \mathbf{W})$ 表示像素 $\mathbf{x}_i$ 的特征， γ 和 λ 为超参数，本

实验中设置为： $\gamma = 2.5$ ， $\lambda = 5$ 。 $\mathcal{L}_s$ 中的第一项保证正常像素的异常得分小于阈值 γ ，第二项保证异常像素和正常像素的异常得分之差大于阈值 λ 。

2.4 特征与分数之间一致性约束损失

在 2.2 和 2.3 节中，本文引入了伪异常数据来辅助训练异常检测模型，使得异常检测器具有更好的泛化性能，能够在测试阶段识别出未知的异常模式。然而，特征约束和分数约束都是对像素特征逐个进行考量，未考虑到像素特征之间的相关性在图像异常检测任务中也起着至关重要的作用。因此，本小节设计了特征与分数之间的一致性损失约束项，通过使像素特征间的拓扑性与分数之间的拓扑性一致，来捕获像素特征之间的相关性。

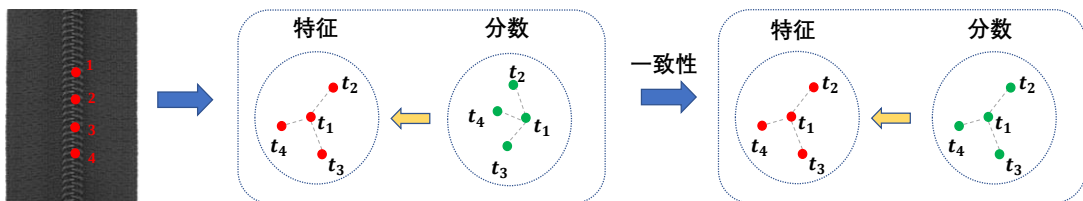


图 6 一致性约束示意图

Fig.6 Consistency constraint diagram

如图 6 所示, 红色点表示像素特征, 绿色点表示像素得分, 不同像素异常得分之间的拓扑结构应与像素特征之间的拓扑结构保持一致, 这里使用 KL 散度^[27]对两种拓扑结构关系进行一致性约束。为了便于后续计算, 本文首先将输入的图片 X 的特征图 $F \in \mathbb{R}^{C_1 \times H \times W}$ 、分数图 $Z \in \mathbb{R}^{1 \times H \times W}$ 的表达形式分别改为 $F \in \mathbb{R}^{K \times C_1}$, $Z \in \mathbb{R}^{K \times 1}$, 其中 $K = H \times W$; 然后分别计算特征图 F 、分数图 Z 中每一个像素的特征向量与所有像素特征向量之间的欧式距离, 得到相应的特征相似度矩阵 $P \in \mathbb{R}^{K \times K}$ 、分数相似性矩阵 $Q \in \mathbb{R}^{K \times K}$, 本文所提出的一致性损失函数定义如下:

$$\mathcal{L}_{fs} = \frac{1}{K} \sum_{i=1}^K \text{KL}(\sigma(Q_i) \| \sigma(P_i)) \quad (6)$$

其中 K 表示图像中的像素点个数, i 表示相似性矩阵的第 i 行, 即 P_i 、 Q_i 分别表示第 i 个像素特征、分数与图像中所有其他像素特征、分数之间的相似性关系, 然后使用 Softmax 函数 σ 将相似性关系转换为概率分布, 最后通过计算每一行概率分布之间的 KL 散度约束特征与分数之间的一致性关系。假设每个 batch 有 N 张图像, 则最终的一致性损失函数定义如下:

$$\mathcal{L}_{\text{batch_fs}} = \frac{1}{N} \sum_{i=1}^N \mathcal{L}_{fs} \quad (7)$$

由于计算所有像素点之间相似度代价十分高昂, 而且真实异常形式复杂多样, 其特征间的相关性难以捕获。因此, 本文在实验中采用将多个像素点打包成像素块(Patch)的方式, 仅计算了正常像素块之间的一致性关系。

2.5 全局损失函数

在本文中, 全局损失函数包括了特征约束损失、分数约束损失和一致性损失, 具体形式如下:

$$\mathcal{L} = \alpha \mathcal{L}_f + \beta \mathcal{L}_s + \xi \mathcal{L}_{\text{batch_fs}} \quad (8)$$

其中 α , β , ξ 为超参数, 本实验中设置为: $\alpha=1$, $\beta=1$, $\xi=1$ 。

2.6 异常分数

在推理阶段, 本文首先使用学习得到的分数映射网络 η 为每个测试对象生成异常分数图, 如公式(9)所示; 然后采用双线性插值技术将分数图上采样为原图大小, 得到像素级异常分数, 用于图像异常定位; 最后, 本文使用每张图像像素中最大的异常分数作为该图像的异常得分。

$$\text{score}(X_i) = \eta(\psi(X_i, W), \theta_s) \quad (9)$$

3 实验

本研究在 MVTEC^[1]数据集上进行实验评估, 该数据集包含 5354 张高分辨率彩色图像, 分为 10 个对象类别和 5 个纹理类别, 用于异常检测任务。数据集的训练集中仅仅包含正常图像, 测试集由正常图像和具有各种类型缺陷的异常图像组成, 同时, 这个数据集还为每张有缺陷的测试图像提供了像素级注释。图 7 展示了该数据集部分类别的一个正常示例图像和一个异常示例图像以及异常所对应的真值图像即 Ground_truth, 为了便于观察, 本文还将异常示例图像中的部分异常区域进行了放大显示。从异常示例图像中可以看出, 测试集中的异常图像所包含的缺陷类型各种各样, 如划痕、凹痕和破损等物体表面的缺陷, 或物体部分结构扭曲的缺陷以及物体不完整而显示出的缺陷。由于该数据集图像数量相对较少, 其中每类训练集的图像数量在 60 到 391 张不等, 分辨率在 700×700 和 1024×1024 之间, 且训练集中不包含异常图像, 这给深度表征学习带来了挑战, 因此本文在训练模型时, 通过复制操作(Repeat)对训练集中的数据进行 10 倍扩充, 并在扩充后选取一半数量的图像用于生成伪异常数据。与现有的方法一致, 我们为 MVTEC 数据上每个类别单独训练一个分类器。

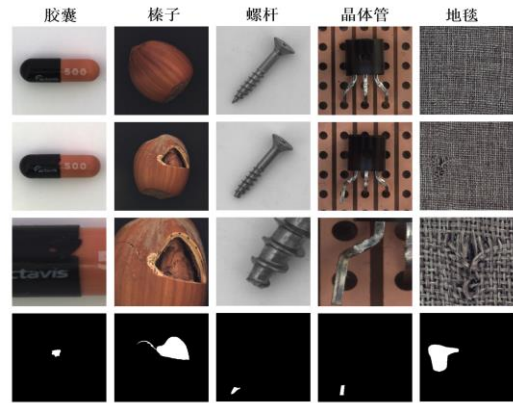


图 7 MVTEC 数据集图像

Fig.7 Image of MVTEC dataset

3.1 实验细节

本文的主干网络为 ResNet50, 参数初始化为在 ImageNet 上预训练参数, 使用 STLNet 纹理网络提取底层纹理特征和自校准模块扩大感受野。采用随机梯度下降(Stochastic Gradient Descent,

SGD)作为优化器。初始学习率设置为 0.001，衰减率为 0.00003，输入图像的分辨率为 256×256，批量大小设置为 32，训练迭代次数 Epoch=200。

3.2 评价指标与实验结果

对于图像级异常检测，本文使用接收机工作特性曲线下面积(AUROC)作为评价指标；对于图像缺陷定位，本文使用像素级 AUROC 和每区域重叠率^[27](Per-Region-Overlap, PRO)作为评价指标，不同于简单地逐像素考量 AUROC 值，PRO 评分对任意大小的异常区域一视同仁。我们将本文的方法与七种近期提出的典型图像异常检测算法进行比较，其中包括 US^[28]，MF^[29]，PSVDD^[30]，FCDD^[21]，PaDiM^[16]，CutPaste^[17]以及 RD^[31]。

本文方法在 MVTEC 数据集上的实验结果如表 1 和表 2 所示，15 个类别图像级检测平均精度和像素级检测平均精度均高于现有方法。图像级平均 AUROC 值达到 98.7%，像素级平均 AUROC 值和 PRO 值分别达到 97.9%和 94.2%。为了研究本文方法对各种异常的稳健性，本文将缺陷类型分类为较大缺陷和微小缺陷两类，并通过图可视化定性评价其性能。如图 8 所示，第一行为 MVTEC 数据集中部分异常图像，第二行为 MVTEC 数据集为异常图像所对应的真值图像，其中白色区域表示图像中具体的异常区域，第三行为模型的可视化结果，可以看出本文的方法在整体上产生了优异的检测效果。

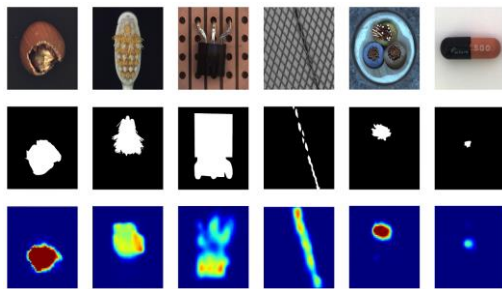


图 8 异常检测可视化图

Fig.8 Anomaly detection visualization

3.3 伪异常样本比例分析

本文通过在训练集中引入伪异常样本，将异常检测任务转化伪监督学习任务，本小节研究不同伪异常样本比例对本文方法的性能影响。表 3 给出了六种伪异常样本比例下本文方法的实验结果，表中 AUROC_{AD} 和 AUROC_{AL} 及 PRO 分别表示 MVTEC 数据集 15 个类别图像级检测的平均精度和像素级检测的平均精度。通过对表 3 中实验

结果进行分析，得到如下发现：(1)本文方法在训练集中不加入伪异常时，性能表现很差。这是因为本文的特征约束损失和分数约束损失，即公式(4)和公式(5)，都是针对训练集中存在异常数据而设计的，此时公式(4)和公式(5)的第二项均会失效，导致实验结果不理想。(2)模型的检测精度随着训练样本中伪异常比例的增加，呈现出先上升再下降的趋势，并且当伪异常样本占有所有训练样本的比例为 50%时，异常检测性能最优。(3)相比于仅用训练集中 5%的数据生成伪异常，本文方法在伪异常样本占有所有训练样本 50%时，图像级检测平均 AUROC 值提升 1.1%，像素级检测平均 AUROC 值和 PRO 值分别提升 0.6%和 3.9%，证实了在训练时加入适量伪异常样本可以显著提升模型性能。(4)模型在 75%和 100%伪异常比例下的实验结果均劣于 50%伪异常比例下的实验结果，这是因为在训练集中引入过多的伪异常样本，会使模型过于关注对伪异常模式的学习，导致模型泛化性能下降。

3.4 消融实验

本小节通过大量消融实验，对自校准模块 STLNet 以及一致性损失的有效性进行验证。本文将多尺度特征融合网络作为基线模型，该网络直接融合 ResNet50 前三层网络生成的特征作为其最终输出，仅使用特征约束训练该网络，直接将像素特征点到超球体球心的距离转化成异常分数。表 4 中的 AUROC_{AD} 和 AUROC_{AL} 及 PRO 分别表示 MVTEC 数据集 15 个类别图像级检测的平均精度和像素级检测的平均精度，可以观察到自校准模块、STLNet、分数约束以及一致性损失都在一定程度上提高了异常检测性能。

4 结论

异常模式通常是罕见的、形式多样的，并且难以获取充分的异常数据用于模型训练，这为异常检测任务带来了极大的挑战。因此，本文首先通过自监督的方式生成伪异常数据，扩充了数据集；然后构建一个多层次特征融合网络，使用 STLNet 抽取底层纹理特征和自校准模块提取更具表示力的特征用于异常检测；最后使用特征约束损失、分数约束损失和一致性损失联合训练网络模型。实验表明本文方法在异常检测和异常定位等方面的性能显著地优于现有的异常检测算法。

表 1 不同方法在 MVTec 数据集的图像级检测结果 AUROC 值(%)

Category/Method	US[28]	MF[29]	PSVDD[30]	PaDiM[16]	CutPaste[17]	RD[31]	Ours
Bottle	99.0	99.1	98.6	99.9	98.2	100	100
Cable	86.2	97.1	90.3	92.7	81.2	95.0	97.0
Capsule	86.1	87.5	76.7	91.3	98.2	96.3	96.6
Carpet	91.6	94.0	92.9	99.8	93.9	98.9	100
Grid	81.0	85.9	94.6	96.7	100	100	100
Hazelnut	93.1	99.4	92.0	92.0	98.3	99.9	99.8
Leather	88.2	99.2	90.9	100	100	100	100
Metal Nut	82.0	96.2	94.0	98.7	99.9	100	98.6
Pill	87.9	90.1	86.1	93.3	94.9	96.6	96.9
Screw	54.9	97.5	81.3	85.8	88.7	97.0	93.2
Tile	99.1	99.0	97.8	98.1	94.6	99.3	100
Toothbrush	95.3	100	100	96.1	99.4	99.5	100
Transistor	81.8	94.4	91.5	97.4	96.1	96.7	98.9
Wood	97.7	99.2	96.5	99.2	99.1	99.2	99.6
Zipper	91.9	98.6	97.9	90.3	99.9	98.5	100
Average	87.7	95.8	92.1	95.5	96.1	98.5	98.7

表 2 不同方法在 MVTec 数据集的实验结果像素级 AUROC 值(左%)和 PRO 值(右%)

Tab.2 Experimental results of different methods in MVTec data sets at pixel level AUROC value (left %) and PRO value (right %)

Category/Method	US[28]	MF[29]	FCDD[21]	PaDiM[16]	CutPaste[17]	RD[31]	Ours
Bottle	-/93.1	-/88.8	87.0/-	98.3/94.8	97.6/-	98.7/96.6	99.0/97.5
Cable	-/81.8	-/93.7	94.0/-	96.7/88.8	90.0/-	97.4/91.0	95.3/91.8
Capsule	-/96.8	-/87.9	94.0/-	98.5/93.5	97.4/-	98.7/95.8	98.5/93.7
Carpet	-/87.9	-/87.8	99.0/-	99.1/96.2	98.3/-	98.9/97.0	99.2/97.4
Grid	-/95.2	-/86.5	94.0/-	97.3/94.6	97.5/-	99.3/97.6	99.3/97.5
Hazelnut	-/96.5	-/88.6	97.0/-	98.2/92.6	97.3/-	98.9/95.5	99.2/98.1
Leather	-/94.5	-/95.9	99.0/-	99.2/97.8	99.5/-	99.4/ 99.1	99.5/98.9
Metal Nut	-/94.2	-/86.9	98.0/-	97.2/85.6	93.1/-	97.3/92.3	98.1/93.1
Pill	-/96.1	-/93.0	95.0/-	95.7/92.7	95.7/-	98.2/96.4	97.6/95.5
Screw	-/94.2	-/95.4	93.0/-	98.5/94.4	96.7/-	99.6/98.2	97.2/89.6
Tile	-/94.6	-/88.1	97.0/-	94.1/86.0	90.5/-	95.6/90.6	97.5/94.5
Toothbrush	-/93.3	-/87.7	89.0/-	98.8/93.1	98.1/-	99.1/94.5	98.8/90.9
Transistor	-/66.0	-/92.6	82.0/-	97.5/84.5	93.0/-	92.5/78.0	91.6/80.8
Wood	-/91.1	-/84.8	95.0/-	94.9/91.9	95.5/-	95.3/90.9	97.9/96.8
Zipper	-/95.1	-/93.6	97.0/-	98.5/95.9	99.3/-	98.2/95.4	99.1/ 97.1
Average	-/91.4	-/90.1	94.0/-	97.5/92.1	96.0/-	97.8/93.9	97.9/94.2

表 3 本文方法在六种不同伪异常比例下的实验结果(%)

Metric	0	0.05	0.25	0.5	0.75	1.0
	伪异常	伪异常	伪异常	伪异常	伪异常	伪异常
AUROC _{AD} (%)	68.2	97.6	98.2	98.7	98.4	98.0
AUROC _{AL} (%)	61.4	97.3	97.7	97.9	97.7	97.0
PRO(%)	32.2	90.3	93.5	94.2	93.8	93.0

表 4 消融实验结果(%)

Tab.4 Results of ablation experiment(%)

Metric	基线	基线+ 自校准	基线+自校准 +STL	基线+自校准 +STL+分数约束	基线+自校准 +STL+分数约束+一致性
AUROC _{AD} (%)	97.84	98.3	98.5	98.6	98.7
AUROC _{AL} (%)	97.37	97.7	97.8	97.7	97.9
PRO(%)	92.33	93.2	93.3	94.0	94.2

参 考 文 献

- [1] BERGMANN P, FAUSER M, SATTLEGGER D, et al. MVTec AD--A comprehensive real-world dataset for unsupervised anomaly detection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 9592-9600.
- [2] GUO X, GICHOYA J W, PURKAYASTHA S, et al. Margin-aware intraclass novelty identification for medical images[J]. Journal of Medical Imaging, 2022, 9(1): 014004. doi: 10.1117/1.JMI.9.1.014004.
- [3] SCHLEGL T, SEEBÖCK P, WALDSTEIN S M, et al. Unsupervised anomaly detection with generative adversarial networks to guide marker discovery[C]. The 25th International Conference on Information Processing in Medical Imaging, Springer, Cham, 2017: 146-157. doi: 10.1007/978-3-319-59050-9_12.
- [4] LIU W, LUO W, LIAN D, et al. Future frame prediction for anomaly detection--a new baseline[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 6536-6545.
- [5] LIU Y, Li C L, PÓCZOS B. Classifier Two Sample Test for Video Anomaly Detections[C]//BMVC. 2018: 71.
- [6] SULTANI W, CHEN C, SHAH M. Real-world anomaly detection in surveillance videos[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 6479-6488.
- [7] SCHRYNEMACKERS M, WEHENKEL L, BABU M M, et al. Classifying pairs with trees for supervised biological network inference[J]. Molecular BioSystems, 2015, 11(8): 2116-2125. doi: 10.1039/C5MB00174A.
- [8] SUN H, ZHENG X, LU X. A supervised segmentation network for hyperspectral image classification[J]. IEEE Transactions on Image Processing, 2021, 30: 2810-2825. doi: 10.1109/TIP.2021.3055613.
- [9] PANG G, SHEN C, CAO L, et al. Deep learning for anomaly detection: A review[J]. ACM Computing Surveys, 2021, 54(2): 1-38. doi: 10.1145/3439950.
- [10] GONG D, LIU L, LE V, et al. Memorizing normality to detect anomaly: Memory-augmented deep autoencoder for unsupervised anomaly detection[C] //Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019: 1705-1714.
- [11] HOU J, ZHANG Y, ZHONG Q, et al. Divide-and-assemble: Learning block-wise memory for unsupervised anomaly detection[C] //Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021: 8791-8800.
- [12] PARK H, NOH J, HAM B. Learning memory-guided normality for anomaly detection[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020: 14372-14381.
- [13] ZAVRTANIK V, KRISTAN M, SKOČAJ D. Reconstruction by inpainting for visual anomaly detection[J]. Pattern Recognition, 2021, 112: 107706. doi: 10.1016/j.patcog.2020.107706.
- [14] ZONG B, SONG Q, MIN M R, et al. Deep autoencoding gaussian mixture model for unsupervised anomaly detection[C]//International Conference on Learning Representations. 2018.
- [15] GENC E U, AHUJA N, NDIOUR I J, et al. Energy-Based Anomaly Detection and Localization[C]. The 9th International Conference on Learning Representations, Vienna, Austria, 2021.
- [16] DEFARD T, SETKOV A, LOESCH A, et al. Padim: a patch distribution modeling framework for anomaly detection and localization[C]. The 27th International Conference on Pattern Recognition, Springer, Cham, 2021: 475-489. doi: 10.1007/978-3-030-68799-1_35.
- [17] LI C L, SOHN K, YOON J, et al. Cutpaste: Self-

- supervised learning for anomaly detection and localization[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021: 9664-9674.
- [18] REN J, LIU P J, FERTIG E, et al. Likelihood ratios for out-of-distribution detection[C]. The 32nd International Conference on Neural Information Processing Systems, Vancouver Canada, 2019, 32.
- [19] RUFF L, VANDERMEULEN R, GOERNITZ N, et al. Deep one-class classification[C]. The 35th International Conference on Machine Learning, Stockholm, Sweden, 2018: 4393-4402.
- [20] RUFF L, VANDERMEULEN R A, GÖRNITZ N, et al. Deep semi-supervised anomaly detection[C]. The 8th International Conference on Learning Representations, Addis Ababa, Ethiopia, 2020.
- [21] LIZNERSKI P, RUFF L, VANDERMEULEN R A, et al. Explainable deep one-class classification[C]. The 9th International Conference on Learning Representations, Vienna, Austria, 2021.
- [22] RUFF L, VANDERMEULEN R A, FRANKS B J, et al. Rethinking assumptions in deep anomaly detection[C]. The 38th International Conference on Machine Learning, Virtual Conference, 2021.
- [23] HENDRYCKS D, MAZEIKA M, DIETTERICH T. Deep anomaly detection with outlier exposure[C]. The 7th International Conference on Learning Representations, Orleans, France, 2019.
- [24] ZHU L, JI D, ZHU S, et al. Learning statistical texture for semantic segmentation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021: 12537-12546.
- [25] LIU J J, HOU Q, CHENG M M, et al. Improving convolutional networks with self-calibrated convolutions[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020: 10096-10105.
- [26] 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.
- [27] YANG C, ZHOU H, AN Z, et al. Cross-image relational knowledge distillation for semantic segmentation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022: 12319-12328.
- [28] BERGMANN P, FAUSER M, SATTLEGGER D, et al. Uninformed students: Student-teacher anomaly detection with discriminative latent embeddings[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020: 4183-4192.
- [29] WU J C, CHEN D J, FUH C S, et al. Learning unsupervised metaformer for anomaly detection[C] //Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021: 4369-4378.
- [30] YI J, YOON S. Patch svdd: Patch-level svdd for anomaly detection and segmentation[C] //Proceedings of the Asian Conference on Computer Vision. 2020.
- [31] DENG H, LI X. Anomaly Detection via Reverse Distillation from One-Class Embedding[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022: 9737-9746.