

基于离散-连续特征耦合的图像异常检测算法

刘洋, 侯春萍, 葛棒棒, 王致芃*, 彭程

天津大学电气自动化与信息工程学院, 天津 300072

摘要 光学图像异常检测旨在仅利用正常样本训练模型, 并检测出偏离正常规律的异常样本。为了解决生成式异常检测算法中的普适性重建和低质量干扰问题, 基于自编码网络, 提出一种新的图像异常检测算法。首先将隐空间特征变换为连续特征和离散特征, 即块描述特征和块哈希特征, 哈希特征具有二值化特性, 可以避免隐空间欠采样, 进而解决普适性重建问题; 其次基于离散-连续特征的耦合关系, 利用图收缩方法建立块相似性矩阵构建起哈希特征和描述特征间的关联, 并通过该关联提出一种块间重构方法, 保证对原始图像的高质量重构, 从而解决低质量干扰问题。在国际公开数据集 MVTEC AD 上进行了验证实验, 实验结果表明, 所提算法的精度优于当前同类型的异常检测算法。

关键词 机器视觉; 光学图像; 异常检测; 哈希特征

中图分类号 TP391.4

文献标志码 A

doi: 10.3788/LOP202259.0815009

Image Anomaly Detection Algorithm Based on Discrete-Continuous Feature Coupling

Liu Yang, Hou Chunping, Ge Bangbang, Wang Zhipeng*, Peng Cheng

School of Electrical and Information Engineering, Tianjin University, Tianjin 300072, China

Abstract The purpose of optical image anomaly detection is to train the model only with normal samples and detect abnormal samples that deviate from the normal law. To solve the universal reconstruction and low-quality interference problems in the generation-based anomaly detection algorithm, a new image anomaly detection algorithm is proposed based on the autoencoder network. First, the latent features are transformed into continuous and discrete features, namely block descriptive and hash features. Hash features have binarization characteristics; it can avoid under-sampling of latent space, thereby the problem of universal reconstruction can be effectively solved. Second, Based on the coupling relationship of discrete-continuous features, the graph shrinkage method is used to establish the block similarity matrix which constructs the association between hash and description features. Then the interblock reconstruction method is proposed to ensure high-quality reconstruction of the image and solving the problem of low-quality interference. Experiments on the international public dataset, MVTEC AD, prove that the accuracy of the proposed algorithm is better than the present anomaly detection algorithms.

Key words machine vision; optical image; anomaly detection; Hash feature

收稿日期: 2021-07-06; 修回日期: 2021-08-15; 录用日期: 2021-08-25

基金项目: 国家自然科学基金(61731003)、西藏自治区科技计划项目(XZ201901-GB-16)、赛尔网络下一代互联网技术创新项目(NGII20180115)

通信作者: *zpwang@tju.edu.cn

1 引言

现阶段,人工智能技术给光学图像分析带来巨大变革,对现代医疗^[1-2]、制造业^[3]及光谱遥感^[4-5]等领域产生了深远影响。在光学图像智能分析过程中,受病灶、缺陷和成像传感器响应畸变等因素的影响,异常样本不可避免,无法筛选出该类样本会对医学图像分割^[6]、产品质量保障^[7]和遥感目标分析^[8]造成不可估计的损失。异常样本外观多样、分布不一致且数量极少,因此传统的监督分类算法和成熟的深度神经网络难以适用于异常检测任务^[9-12]。而无监督图像异常检测旨在仅利用正常样本训练的情况下构建模型,并检测出偏离正常规律的异常样本,具有重要的理论意义和应用价值。

目前,无监督异常检测算法有判别法和重构法两种。判别法旨在利用某种假设构造异常的样本训练精准分类器,从而检测出真正异常样本。常见的假设包括:伪异常样本^[13]、偏离正常空间样本^[14]等。这类算法的优点是计算量小,但通常情况下仅适用于全局异常样本,无法对局部异常进行定位,导致其适用场景具有局限性。

重构法旨在构建可以重构原始图像的网络模型,由于训练集仅为正常样本,因此模型仅学得正常样本的外观分布规律。在测试阶段,模型仅可以实现对正常样本的清晰重构,而对异常样本会产生较大的重构误差^[15]。近年来,基于自编码器(AE)和生成对抗网络(GAN)的重构法取得了很大进步。这些方法可以分为两类:1) 基于正常样本的高质量重构,利用清晰化重构模型使正常区域获得更小的重构误差,则异常区域重构误差相对更大。Akçay等^[16]提出了GANomaly,该方法利用对抗自编码网络优化了对隐空间的搜索过程,直接以重构清晰的原始图作为训练目标,获得了较大的性能提升。随后,Akçay等^[17]又引入了跳层连接方式进一步提升重构图像的质量。2) 基于异常样本的低质量重构,引入隐空间限定表达机制来训练正常样本,因异常样本无法在限定空间上表达,则异常区域重构误差更大。Gong等^[18]在隐空间中引入字典结构,使输入解码器的特征向量限制表达在训练好的字典上,对于异常样本而言,由于用于解码的特征是正常特征,理想情况下会和原始输入图像产生较大的重构误差。Park等^[19]优化了这个模型的训练过程并应用到视频异常事件检测任务中。

上述方法在一定程度上推动了异常检测算法的进步,但两类方法都存在局限性,主要有以下两点原因。1) 普适性重建问题,高质量重构方法需要信息丰富的浅层特征,而浅层特征具有通用性,因此在训练阶段存在隐空间欠采样现象,导致一些情况下的异常样本也可以获得良好的重构效果;2) 低质量干扰问题,限定表达机制限制了特征的多样性,降低了模型对差异性内容的重构,会导致某些正常样本的重构质量差,降低了算法的区分度。

为了解决这些问题,本文基于离散-连续特征的耦合关系提出了一种新的图像异常检测算法,该算法融合了这两类方法的优点,可以实现正常区域高质量重构和异常区域低质量重构,规避了普适性重建问题和低质量干扰问题。所提算法融合了哈希特征和描述特征的优点,利用哈希特征的有限编码特性,避免了隐空间欠采样现象,从而解决一般描述特征的普适性重建问题。以往的研究主要基于样本相似性来构建模型,然而正常样本多属于同类物体,具有稳定的块间相似性,而该特性在以往的研究中并没有得到充分利用。所提算法利用离散-连续特征的耦合关系,利用图收缩方法来构建哈希特征和描述特征间的关联,对汉明距离相近块的描述特征进行融合以实现块间重构,进而高质量地重构输入图像,从而解决离散特征的低质量干扰问题。

2 基于离散-连续特征耦合的图像异常检测算法基本原理

所提基于离散-连续特征耦合的图像异常检测算法通过对原始图像进行编码获得隐空间特征,并将该特征变换得到块哈希特征和描述特征,利用哈希特征构建特征间的相似性图,从而生成融合描述特征实现块间重构,进而实现对原始图像的高质量重构。所提算法的整体框架如图1所示,主要包括主干网络(编码网络,解码网络)、特征提取模块和描述特征融合模块3个组成部分。

2.1 编解码器及网络结构

给定正常样本训练集 X ,编码网络的作用是提取隐空间特征,即将图像 $x_i \in X$ 输入编码网络获得隐空间特征 z ;解码网络的作用是将融合描述特征 \hat{z} 解码为原始图像。

$$z = f_e(x; \theta_e), \quad (1)$$

$$\hat{x} = f_d(\hat{z}; \theta_d), \quad (2)$$

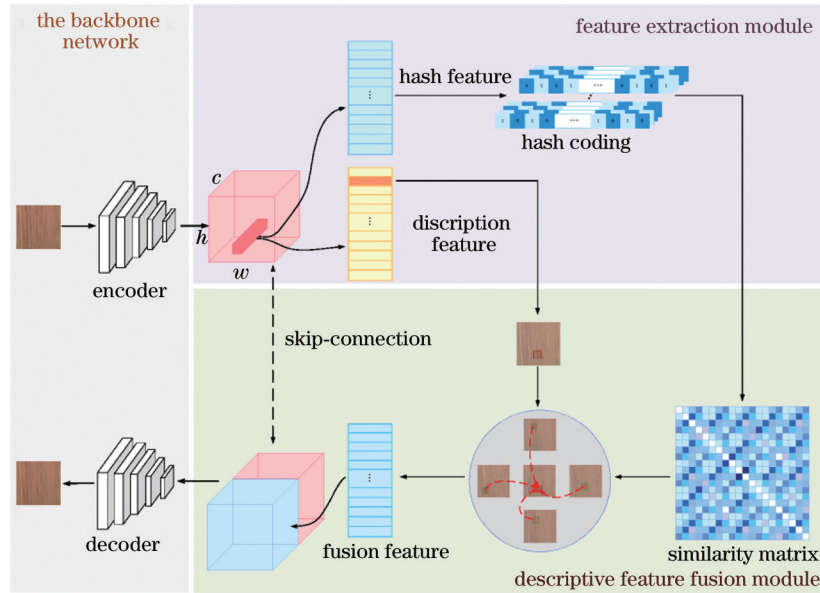


图 1 整体网络训练流程图

Fig. 1 Flow chart of overall network

式中： θ_e 为编码网络 $f_e(\cdot)$ 的网络参数； θ_d 为解码网络 $f_d(\cdot)$ 的网络参数； \hat{x} 为解码网络输出的重构图像。区别于一般的自编码结构，解码网络的输入与编码网络的输出为同一变量 z ，所提算法采用的是融合描述特征 \hat{z} 。

所提算法采用的是生成对抗训练模式，近年来该模式在异常检测任务上具有良好表现。因此，所提算法利用上述的 $f_e(\cdot)$ 和 $f_d(\cdot)$ 组成生成器 $G(\cdot)$ ，并额外引入判别器 $D(\cdot)$ ，共同构成生成对抗网络。生成器以最真实化的重构输入样本 x_i 为目标，判别器以准确判别输入图像为原始图像还是重构图像为学习目标。两者相互促进构成了对抗机制，随着迭代次数的增加，生成器可以生成更加逼真的重构图像，从而使判别器无法感知出重构图像与原始图像的差别，判别器则以越来越灵敏的判别为优化方向。因此，该网络具有两个训练目标：1)内容的真实性，图像间应具有很强的相似性，即原始图像和重构图像像素间的距离很小；2)分布一致性，因网络采用对抗训练方式，当优化完成时，生成图像与原始图像应是概率同分布。因此，整个网络的训练损失由两部分构成。

$$\mathcal{L}_m = E_{x \in X} \|x - \hat{x}\|_2, \quad (3)$$

$$\mathcal{L}_{adv} = E_{x \in X} [\log D(x)] + E_{x \in X} \{\log [1 - D(\hat{x})]\}, \quad (4)$$

式中： \mathcal{L}_m 是重构损失； \mathcal{L}_{adv} 是对抗损失； $\|\cdot\|_2$ 表示图像级的欧氏距离； $E(\cdot)$ 表示分布函数的期望值，对抗

损失采用交叉熵作为损失函数。生成器通过最小化重构损失 $\min_G \mathcal{L}_m$ 和对抗损失 $\min_G \mathcal{L}_{adv}$ ，判别器通过最大化对抗损失 $\max_D \mathcal{L}_{adv}$ 以达到训练目标。

2.2 哈希隐空间和描述隐空间特征提取

生成式异常检测方法的基本原理是利用(1)式提取编码特征，并在训练过程中直接使用该特征对原始图像进行清晰化重构。由于训练样本仅为正常样本，网络只具备准确描述正常样本的能力。当输入为异常样本时，因其局部外观或整体结构在训练阶段中没有出现过，理想情况下，网络无法对该区域或该样本进行高质量重构，造成重构出的图像质量差，与输入图像之间存在较大的重构误差，进而可以通过设定重构误差阈值的方式来区分正常与异常样本。

目前，该类方法虽然取得了一定的效果，但是也存在着明显的不足，致使该方法无法达到理想效果，影响了该类方法的实际应用。1) 该类方法有效的前提是对正常样本进行高质量重构，因此编码器多采用浅层网络以避免细节信息的丢失。并且网络仅以重构原始图像为优化目标，过程中没有高级语义信息的参与，这导致所获得的隐空间特征为纹理、颜色等浅层特征，然而在自然场景下浅层特征具有通用性，对应的生成网络也存在过表达现象，当异常样本的浅层信息与正常样本较为相似的情况下，异常样本也可以具有较好的重构效果。2) 该类方法的另一个假设是隐空间采样假设，即确保隐

空间的所有特征都可以重构出正常样本,然而高维隐空间范围过大,枚举隐空间特征不切实际,欠采样现象不可避免。由于哈希特征具有良好的枚举性,Chen 等^[20]曾尝试采用哈希编码作为隐空间特征来解决欠采样问题。然而,哈希特征存在信息丢失,致使特征的描述能力下降的问题,直接使用哈希特征进行重构无法保证重构图像的质量,影响正常样本与异常样本的区分度,该方法的适用场景受限,并不适用于一般的光学影像分析任务。

为了解决上述问题,所提算法联合哈希特征和描述特征的优点,即哈希特征的二值化特性和描述特征的细节表达能力,可以防止欠采样现象发生从而对原始图像进行清晰重构。特征提取过程如下,给定一张输入图像的编码特征 $\mathbf{z} \in \mathbf{R}^{W \times H \times C}$,其中 W, H, C 分别代表隐空间特征的长度、宽度和通道数量,将该特征按空间展开获得块描述特征 $\mathbf{z}_i \in \mathbf{R}^{1 \times C}, i = 1, \dots, N$,其中特征数量 $N = W \times H$ 。基于该描述特征,通过二值化映射函数,可以获得与之对应哈希特征 \mathbf{b} 。

$$\mathbf{b}_i = \alpha [f_h(\mathbf{z}_i; \theta_h), \epsilon], \quad (5)$$

式中:对应的 M 维的哈希特征 $\mathbf{b}_i \in \{0, 1\}^M; f_h(\cdot)$ 表示特征变化网络, θ_h 为其网络参数,采用 Sigmoid 激活函数层,可以将描述特征映射至 0, 1 二值空间从而生成对应的哈希特征。由于二值化的约束,无法直接使用反向传播方法对哈希特征进行优化,所提算法采用文献[21]的方法加入了离散统计激活层 $\alpha(\cdot, \epsilon)$,该层采用一组随机变量 $\epsilon \sim \mu(0, 1)^M$ 。

$$\mathbf{b}_i^j = \alpha [f_h(\mathbf{d}_i; \theta_h), \epsilon^j] = \begin{cases} 1, & f_h(\mathbf{d}_i; \theta_h)^j \geq \epsilon^j \\ 0, & f_h(\mathbf{d}_i; \theta_h)^j < \epsilon^j \end{cases}, \quad (6)$$

式中: j 代表的第 j 位哈希码; \mathbf{d}_i 是由 \mathbf{z}_i 通过全连接层降维得到的。加入该激活层后,反向传播能以分布导数的方式通过离散编码,进而可以优化哈希特征。同时,为了保证哈希码的信息量最大化,需要保证哈希特征在统计上具有相等数量的 0 和 1,因此哈希特征应服从二项分布。对此,所提算法引入了额外的判别器 $D_H(\cdot, \theta_{D_H})$,用以分类所提取到的哈希特征与二项分布向量。

$$\begin{cases} D_H(\mathbf{b}, \theta_{D_H}) \in (0, 1) \\ D_H(\mathbf{s}, \theta_{D_H}) \in (0, 1) \end{cases}, \quad (7)$$

式中: $\mathbf{s} \in (M, 0.5)$ 是符合二项分布的随机向量。优化哈希特征 \mathbf{b} 的目标是希望哈希特征与二项分布向量概率同分布,其损失函数为

$$\mathcal{L}_{\text{HadV}} = \mathbb{E}_{\mathbf{s} \in (M, 0.5)} [\log D_H(\mathbf{s})] + \mathbb{E}_{\mathbf{b} \in \mathbf{H}} \{\log [1 - D_H(\mathbf{b})]\}, \quad (8)$$

式中: \mathbf{H} 代表哈希特征空间。

2.3 基于汉明相似性的描述特征融合

前文描述了如何提取描述特征与哈希特征,本小节主要研究如何利用哈希特征和描述特征间的耦合关系建立特征间的关联,进而避免直接利用描述特征对原始图像进行重构造成的普适性重建问题及直接使用哈希特征造成的重构误差大、区分能力弱的问题。

对此,所提算法建立了基于汉明距离的块相似图矩阵,由于哈希特征是由描述特征变化得到的,因此哈希特征相似的对应描述特征也应该相似,并且在正常情况下由于外观的有限性,相似的描述特征应具备互表达特性。为了实现这一目标,所提算法采用可优化的图矩阵,根据重建质量引导块间相似性的优化,块间重构关系越强则重构质量高,对应块的相似度更高、哈希特征的汉明距离更小。

对于正常样本而言,样本间具有统一的外观及纹理结构,因此具备明显的块间重构特性,待重构块可以利用相似块的描述特征来清晰化重构外观及结构信息即特征间的互表达特性。对此,所提算法利用相似性矩阵选择部分最相似块,并用对应的描述特征生成融合描述特征,将融合描述特征输入解码器用以重构原始图像。所提算法采用融合描述特征进行重构,可以避免直接使用哈希特征带来的重构图像质量差的问题,也可以避免直接利用块描述特征造成的普适性重建问题。对于异常图像,虽然其描述特征是未见特征,但因哈希特征具有枚举性,对应的哈希特征依然会落入正常样本的哈希特征空间,进而基于该哈希特征寻找到的最近邻块为正常样本,对应的融合描述特征重构出的图像会自然偏向正常样本的外观,从而与原始图像产生较大的重构误差,如图 2 所示,其中图 2(a) 表示异常图像的相似块选择过程,图 2(b) 表示基于相似块的块间重构过程。

具体而言,假设批大小为 n ,则描述特征矩阵 $\mathbf{Z} = [\mathbf{z}_1; \mathbf{z}_2; \dots; \mathbf{z}_N] \in \mathbf{R}^{N \times C}$,对应哈希特征的矩阵 $\mathbf{B} = [\mathbf{b}_1; \mathbf{b}_2; \dots; \mathbf{b}_N] \in \{0, 1\}^{N \times M}$,其中 $N = n \times W \times H$,向量下标代表块索引值。利用哈希特征矩阵 \mathbf{B} 构建块相似矩阵,其中每一个顶点代表一个块,定点间连接线代表块与块的相似度,相似度越大则汉明距离越小,相似度越小则汉明距离越大。因此,

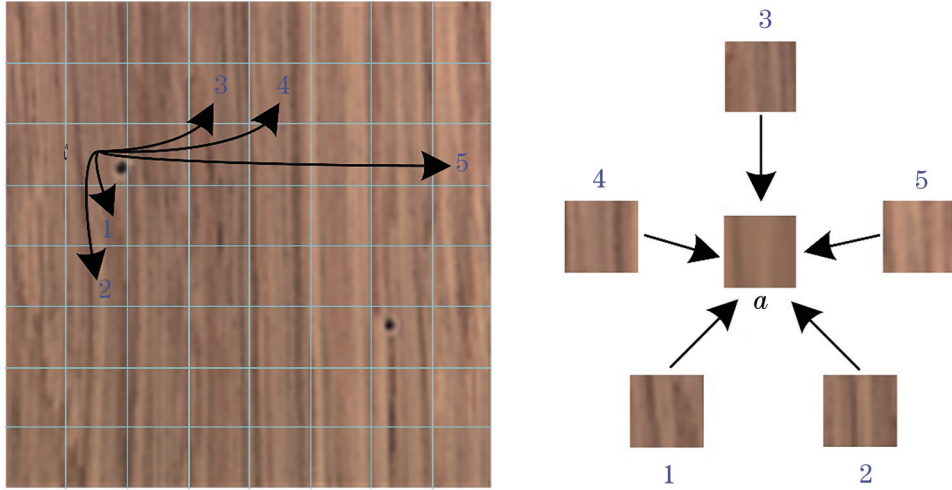


图 2 相似块选择示意图。(a)块选择;(b)块重构

Fig. 2 Schematic diagram of similar block selection. (a) Block selection; (b) block reconstruction

归一化的相似度矩阵的表达式为

$$\mathbf{S} = \mathbf{J} + \frac{1}{M} [\mathbf{B}(\mathbf{B} - \mathbf{J})^T + (\mathbf{B} - \mathbf{J})\mathbf{B}^T], \quad (9)$$

式中: \mathbf{J} 代表全 1 矩阵; \mathbf{S} 是相似度矩阵, 其中每一个元素 $S_{i,j}$ 表示块 i 与块 j 的相似度, $S_{i,j} = 1 - \text{Hamming}(\mathbf{b}_i, \mathbf{b}_j) / M$, $\text{Hamming}(\mathbf{b}_i, \mathbf{b}_j)$ 表示之间的汉明距离。

并且由于正常样本通常是同类物体, 多数块间都具备一定的相似性, 如果使用所有相似块进行块间重构, 会造成某些异常块也会被高质量重构。因此, 在构建相似图的过程中, 所提算法提出一种隐空间相似矩阵的图收缩方法, 即仅保留相似度高于阈值 α 的边, 其余边不予保留, 即

$$\widehat{S}_{i,j} = \begin{cases} S_{i,j}, & S_{i,j} > \alpha \\ 0, & S_{i,j} \leq \alpha \end{cases} \quad (10)$$

网络训练的目的是实现块间重构, 即汉明距离近的块对应得到的描述特征可以互相融合转换。为实现这一目的, 所提算法采用图卷积层方式。

$$\mathbf{Z}' = \text{Sigmoid}(\mathbf{D}^{-\frac{1}{2}} \widehat{\mathbf{S}} \mathbf{D}^{-\frac{1}{2}} \mathbf{Z}_B \mathbf{W}_\theta), \quad (11)$$

式中: $\mathbf{W}_\theta \in \mathbf{R}^{C \times C}$ 为可训练的变换参数; \mathbf{D} 为对角矩阵, $D_{ii} = \sum_j \widehat{S}_{i,j}$ 。将变换后得到的融合描述特征 \mathbf{Z}' 输入解码器 $f_d(\cdot)$ 重建整张图像。当完成训练目标后, 块间重构性强的块具有较小的汉明距离, 所对应的融合描述特征可以重构原始图像。

2.4 网络训练及异常检测

为了保证重构图像的内容真实性、与原始样本的对抗性、生成哈希码的均衡性, 网络的总体损失函数表达式为

$$\mathcal{L} = \mathcal{L}_m + \mathcal{L}_{adv} + \mathcal{L}_{Hadv} \quad (12)$$

在测试过程中, 通过计算原始图像和重构图像的重构损失 \mathcal{L}_m 来获得初始异常分数, 进而进行全局归一化并生成最后的异常分数。

$$S'_i = \frac{S_i - \min(S)}{\max(S) - \min(S)}, \quad (13)$$

式中: S'_i 表示第 i 张图像对应的异常分数; S 表示每张测试图像对应的最大重构误差形成的集合, $\min(S)$ 和 $\max(S)$ 分别代表集合中的最小值和最大值; S_i 表示第 i 张图像对应的最大重构误差。

3 实验与结果分析

3.1 实验环境及实验数据集

实验采用 CPU 为 Intel E-2124、主频为 3.3 GHz、单块 NVIDIA 2080TI 显卡的图形工作站, 操作系统为 Ubuntu 16.04, 采用 Pytorch 1.4 作为深度学习框架。在训练过程中, 采用 step 学习策略, 初始学习率设为 5×10^{-4} , batch size 设为 24, epoch 设置为 200。

为验证所提算法的有效性, 训练及测试数据采用国际公开的 MVTec AD 数据集^[22]。该数据集由 MVTec 提出, 为采集于工业现场的真实缺陷样本图像, 包含 15 类情况并具有纹理和外观两类异常图像, 且每一类都具有训练集和测试集, 训练集全部为该类正常样本图像, 测试集中的数据中混合了正常与异常数据且异常数据具有多种类型。在实验过程中, 将输入图像的分辨率统一设置为 128 pixel \times 128 pixel。在当前的实验环境及数据集设置下, 判断一张分辨率为 128 pixel \times 128 pixel 的图像类型的平均计算时间为 0.0375 s。

3.2 评估指标及对比算法

在异常检测领域, receiver operating characteristic curve (ROC) 能够准确度量检测性能的优劣, 其中 ROC 曲线下的面积被定义为 area under curve (AUC), 取值范围为 $[0, 1]$, 该数值可以有效描述算法对正样本与负样本的区分能力, 值越大表明算法性能越好, 是异常检测领域最为通用的评价指标。因此, 本实验组采用图像级 AUC 对算法进行性能评估。

为了验证所提算法的有效性和先进性, 本实验组对比了近年来异常检测领域的多种优秀算法, 包括: AVID^[23]、AE-L2^[24]、AnoGAN^[25]、DSVDD^[26]、VAE-grad^[27]、GT^[28]、GANomaly^[16] 和 ARNet^[29]。AE-L2 是利用自编码网络并以像素级二范数进行优化的算法, 是目前异常检测领域的性能基准; AnoGAN 和 GANomaly 采用生成对抗方式, 利用不同隐空间特征搜索方式进行异常检测; AVID 和

DSVDD 强化了判别器对于异常检测的作用; ARNet 是最近提出的一项自监督异常检测工作, 它以图像属性修复作为自监督任务, 以恢复后的图像质量作为判别正常与异常的指标, 并取得了很好的效果。

3.3 客观分析

表 1 为所提算法与对比算法在 MVTec AD 数据集上的评估结果, 表中数值为 AUC 指标, 加粗表示效果优于其他算法。从表 1 可以看出, 所提算法的平均 AUC 较其他异常检测算法有一定提高, 表明所提算法在 MVTec AD 数据集中实现了较好检测效果, 能够有效判断出多种物体的异常情况。具体而言, 所提算法的平均 AUC 高于基准算法 AE-L2 约 16 个百分点, 高于 ARNet 3 个百分点, 对于纹理类异常和结构类异常均有较好的检测效果, 并在 Bottle、Hazelut 等多个子数据集上取得了最佳的实验效果, 验证了所提算法的有效性和先进性。

表 1 不同算法在各类别上的 AUC

Table 1 AUC of different algorithms in each category

unit: %

Category	AVID ^[23]	AE-L2 ^[24]	AnoGAN ^[25]	DSVDD ^[26]	VAE-grad ^[27]	GT ^[28]	GANomaly ^[16]	ARNet ^[29]	Proposed algorithm
Bottle	88.0	80.0	69.0	86.0	86.0	74.0	89.2	94.1	97.3
Hazelnut	86.0	88.0	50.0	71.0	74.0	33.0	78.5	85.5	92.9
Capsule	85.0	62.0	58.0	69.0	86.0	67.0	73.2	68.1	74.3
Metal Nut	63.0	73.0	50.0	75.0	78.0	82.0	70.0	66.7	68.9
Leather	58.0	44.0	52.0	73.0	71.0	82.0	84.2	86.2	67.3
Pill	86.0	62.0	62.0	77.0	80.0	65.0	74.3	78.6	74.7
Wood	83.0	74.0	68.0	87.0	89.0	48.0	83.4	92.3	99.2
Carpet	70.0	50.0	49.0	54.0	67.0	46.0	69.9	70.6	89.3
Tile	66.0	77.0	51.0	81.0	81.0	54.0	79.4	73.5	96.4
Grid	59.0	78.0	51.0	59.0	83.0	62.0	70.8	88.3	91.3
Cable	64.0	56.0	53.0	71.0	56.0	85.0	75.7	83.2	86.3
Transistor	58.0	71.0	67.0	65.0	70.0	80.0	79.2	84.3	87.9
Toothbrush	73.0	98.0	57.0	70.0	89.0	94.0	65.3	100	97.5
Screw	66.0	69.0	35.0	64.0	71.0	44.0	74.6	100	100
Zipper	84.0	80.0	59.0	74.0	67.0	87.0	74.5	87.6	80.5
Mean	73.0	71.0	55.0	72.0	77.0	67.0	76.2	83.9	86.9

图 3 为网络训练精度的动态变化过程, 图 4 为总体损失函数的变化曲线。在训练迭代的过程中, 网络训练精度逐步升高并趋于稳定, 总体损失值不断减小并逐步趋于零, 表明网络在训练的过程中逐步趋于稳定, 具有较好的异常检测性能。

3.4 主观分析

从 MVTec AD 数据集中的 8 个子集中各选取了 2 个样本进行主观分析, 所提算法的结果如图 5 所示。通过对比输入图像与输出图像的图像质

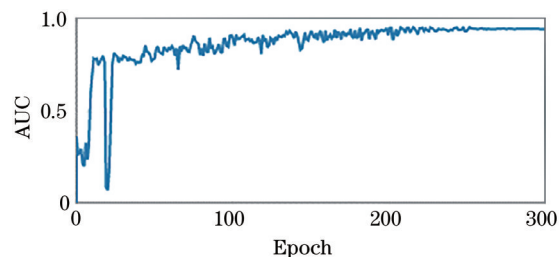


图 3 网络训练精度变化图

Fig. 3 Variation diagram of network training accuracy

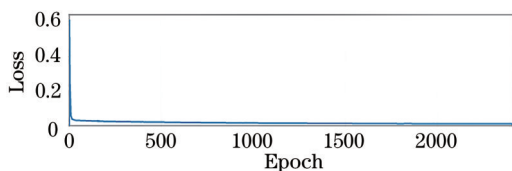


图 4 总体训练损失变化图

Fig. 4 Change diagram of overall training loss

量,可以得到输入图像与输出图像之间的像素级误差。从图 5 可以看出,所提算法的可行性主要来源于以下 3 个方面:1) 对于局部异常可以准确恢复,如 cable、capsule 等类,重构后的异常区域更加偏向正常外观,异常纹理得到良好的修复,重构图像会与原始图像产生较为明显的重构误差。特别是针

对细小缺陷也会有明显的修复效果,例如 capsule、screw 类,表明了所提算法可以有效解决普适性重构问题。2) 对于正常区域的清晰重构,如 bottle、cable 等类,所提算法可以清晰化重构图像的正常区域,特别是对纹理复杂区域有较好的生成效果,可以避免正常区域重建质量差导致的误判,这也表明了所提算法有效解决低质量干扰问题。3) 由 metal 的可视化结果可以看出,当缺陷的外观与正常区域有区分时,所提算法可以有效检测出异常区域;由 tile 的可视化结果可以看出,受网络生成能力的限制,纹理类物体的生成图像存在一定程度的失真,但是由于可以对缺陷区域进行准确的修复,所提算法可以检测出异常位置。

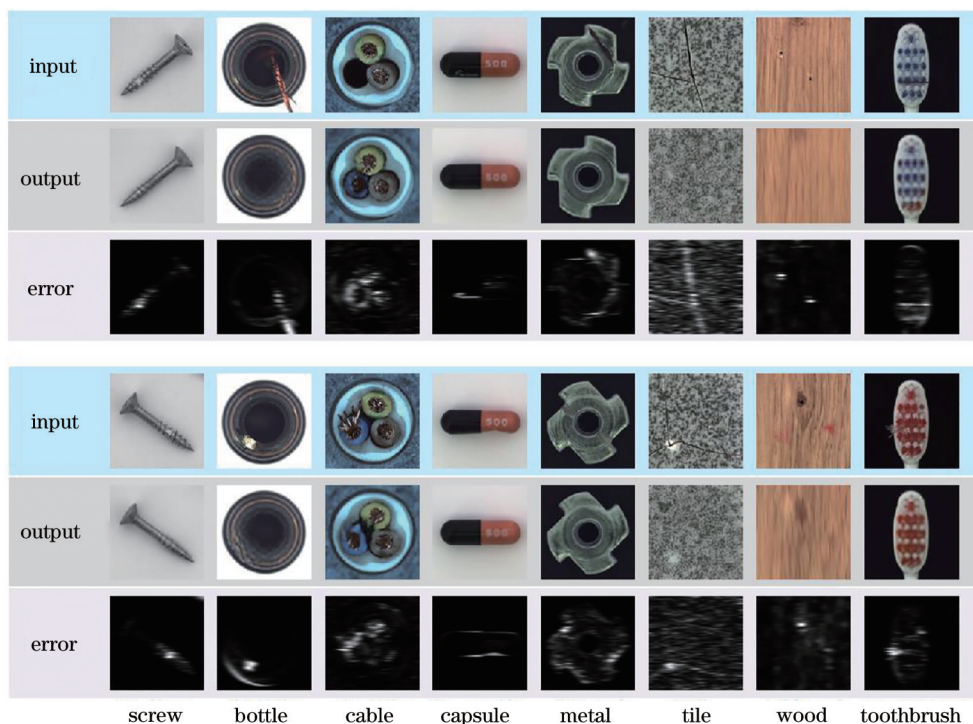


图 5 所提算法对部分异常样本类的检测结果

Fig. 5 Detection results of proposed algorithm for some abnormal sample classes

3.5 消融实验分析

3.5.1 块选择验证实验

为了验证所提算法中块间重构方式的合理性及有效性,设计了可视化块选择验证实验,如图 6 所示。图 6 展示了最近邻块的选择结果图,分别为正常区域和异常区域在汉明空间内的最近邻块,其中[图 6(a)]是 wood 类中的异常样本,[图 6(b)]是 wood 类中的正常样本, X 表示原始图像, \hat{X} 表示重构图像,并根据相似性矩阵选择了与目标块最相似的 5 个块,数字表示了块索引序号。从[图 6(a)]可

以看出,所选择出的相似块为正常块,且纹理结构与目标块较为一致,而[图 6(a)]中的另一异常块并未索引为该目标块的近邻块,这是由于该块的纹理较目标块有较大差别,因此哈希特征间的汉明距离较大、相似性较低。同时,正常块的相似块均为正常块,用这些相似块融合重构出的原始图像可以获得良好的重构质量。这充分说明了所提算法的块间重构方式的合理性及有效性,说明了所提算法可以有效解决低质量干扰问题和普适性重构问题。

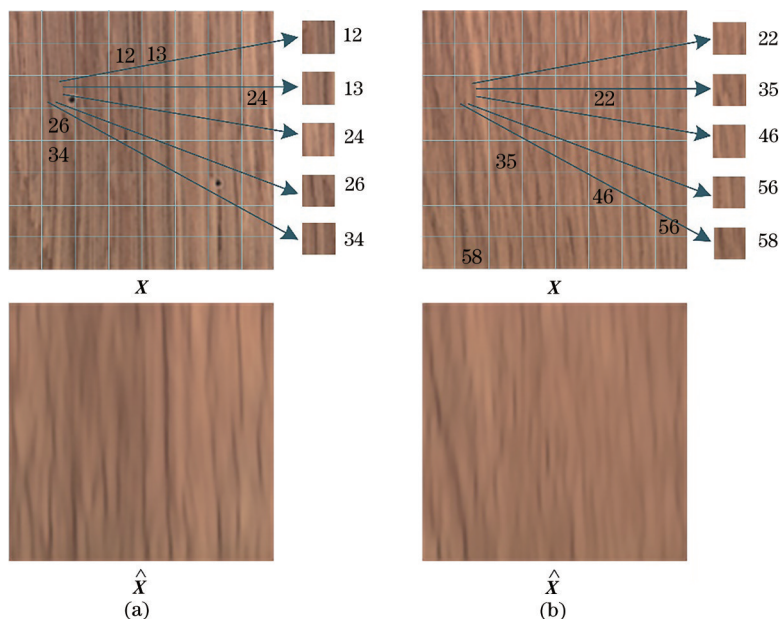


图 6 所提算法在 MVTec AD 数据集上的块选择示意图。(a)异常;(b)正常

Fig. 6 Block selection results of proposed algorithm on MVTec AD dataset. (a) Abnormal; (b) normal

3.5.2 参数验证实验

为了验证块选择阈值 α 选取的合理性及参数的鲁棒性,设计了相关消融实验。在 MVTec AD 数据集的 hazelnut 类上进行实验,结果如图 7 所示,横轴为 α ,左图的取值范围为 $[0, 1]$,右图的取值范围为

$[0.7, 1]$,纵轴表示模型准确率。从图中可以看出,随着阈值的增加,系统性能呈现先增后减的分布趋势,总体变化范围为 1%。实验中,当阈值为 0.95 时获得最优性能。为了保证整个实验的公平性,在实验过程中,所有类的阈值 α 均设置为 0.95。

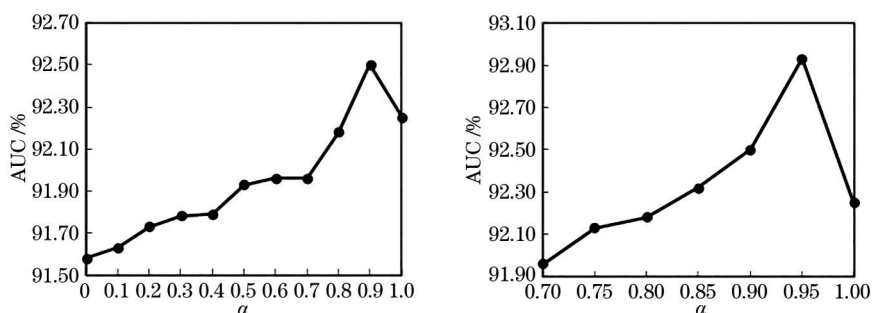


图 7 在 MVTec AD 数据集上,参数 α 与系统性能的关系曲线

Fig. 7 Relationship curves between parameter α and system performance on MVTec AD dataset

3.5.3 初始学习率验证实验

为了验证初始学习率对训练网络整体性能的影响,设计了关于初始学习率的相关消融实验。在 MVTec AD 数据集的 bottle 类上进行实验,结果如图 8 所示,其中横轴表示初始学习率,纵轴表示不同初始学习率对应的网络性能。从图 8 可以看出,当初始学习率在 $[5 \times 10^{-6}, 5 \times 10^{-3}]$ 范围内取值时,网络整体性能较好。为了保证整个实验的公平性,在实验过程中,设置初始学习率为 5×10^{-4} 。

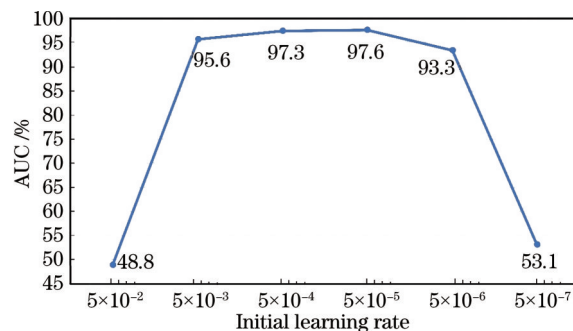


图 8 初始学习率与网络性能的关系曲线

Fig. 8 Relationship curve between initial learning rate and network performance

4 结 论

提出了一种基于离散-连续特征耦合的图像异常检测算法,以自编码器为框架,重点对隐空间的特征变化过程进行优化。该算法利用了离散-连续特征的耦合关系,在隐空间中提取了块哈希特征和块描述特征,并基于正常图像的块间重构特性建立了块相似性矩阵,从而构建了离散-连续特征的关联,生成了融合描述特征,并利用该特征实现了对原始图像的清晰重构。在MVTec AD数据集上进行了验证,并与国际上主流的异常检测算法进行了对比,实验结果表明,所提算法具有更高的图像异常检测准确率。相关的消融实验进一步说明了所提算法的有效性和合理性,验证了所提算法特别是块间重构方式在解决普适性重建问题和低质量干扰问题上的作用。下一步将研究该异常检测算法在遥感图像和医疗图像等其他光学图像智能分析领域的有效性。

参 考 文 献

- [1] Ma R Q, Liao N F, Keizo S. Detection and classification of color deficiency before vision experiment [J]. *Acta Optica Sinica*, 2016, 36(6): 0633001.
马瑞青, 廖宁放, 篠森敬三. 视觉实验前期色觉异常的检测和分类[J]. *光学学报*, 2016, 36(6): 0633001.
- [2] Huang S, Li F F, Chen Q. Computed tomography image classification algorithm based on improved deep residual network[J]. *Acta Optica Sinica*, 2020, 40(3): 0310002.
黄盛, 李菲菲, 陈虬. 基于改进深度残差网络的计算断层扫描图像分类算法[J]. *光学学报*, 2020, 40(3): 0310002.
- [3] Yin S B, Ren Y J, Liu T, et al. Review on application of machine vision in modern automobile manufacturing [J]. *Acta Optica Sinica*, 2018, 38(8): 0815001.
尹仕斌, 任永杰, 刘涛, 等. 机器视觉技术在现代汽车制造中的应用综述[J]. *光学学报*, 2018, 38(8): 0815001.
- [4] Zhao C H, You W, Qi B, et al. Real-time anomaly detection algorithm for hyperspectral remote sensing by using recursive polynomial kernel function[J]. *Acta Optica Sinica*, 2016, 36(2): 0228002.
赵春晖, 尤伟, 齐滨, 等. 采用多项式递归核的高光谱遥感异常实时检测算法[J]. *光学学报*, 2016, 36(2): 0228002.
- [5] Li X F, Wang S Q, Weng X, et al. Remote sensing of floating macroalgae blooms in the East China sea based on UNet deep learning model[J]. *Acta Optica Sinica*, 2021, 41(2): 0201002.
李潇凡, 王胜强, 翁轩, 等. 基于UNet深度学习算法的东海大型漂浮藻类遥感监测[J]. *光学学报*, 2021, 41(2): 0201002.
- [6] Yao H B, Bian J W, Cong J W, et al. Medical image segmentation model based on local sparse shape representation[J]. *Laser & Optoelectronics Progress*, 2018, 55(5): 051011.
姚红兵, 卞锦文, 丛嘉伟, 等. 基于局部稀疏形状表示的医学图像分割模型[J]. *激光与光电子学进展*, 2018, 55(5): 051011.
- [7] Wu Q, Zhang R. Wavelet denoising of near-earth all-day star map based on local outlier factor[J]. *Acta Optica Sinica*, 2020, 40(8): 0810001.
吴强, 张锐. 基于局部异常因子的近地全天时星图小波去噪[J]. *光学学报*, 2020, 40(8): 0810001.
- [8] Ju H H, Liu Z G, Wang Y. Hyperspectral anomaly detection algorithm based on combination of spectral and spatial information[J]. *Laser & Optoelectronics Progress*, 2018, 55(12): 122801.
鞠荟荟, 刘志刚, 汪洋. 基于空谱联合的高光谱异常检测算法[J]. *激光与光电子学进展*, 2018, 55(12): 122801.
- [9] Zhang X H, Hao R F, Li T Y. Hyperspectral abnormal target detection based on low rank and sparse matrix decomposition-sparse representation[J]. *Laser & Optoelectronics Progress*, 2019, 56(4): 042801.
张晓慧, 郝润芳, 李廷鱼. 基于低秩稀疏矩阵分解和稀疏字典表达的高光谱异常目标检测[J]. *激光与光电子学进展*, 2019, 56(4): 042801.
- [10] Zhou P P, Ding Q H, Luo H B, et al. Anomaly detection and location in crowded surveillance videos [J]. *Acta Optica Sinica*, 2018, 38(8): 0815007.
周培培, 丁庆海, 罗海波, 等. 视频监控中的人群异常行为检测与定位[J]. *光学学报*, 2018, 38(8): 0815007.
- [11] Li W T, Fu H, Hao Z, et al. Detection algorithm of crowd abnormal event based on Girvan-Newman splitting[J]. *Laser & Optoelectronics Progress*, 2020, 57(6): 061506.
李文韬, 付晗, 郝真, 等. 基于Girvan-Newman分裂的人群异常事件检测算法[J]. *激光与光电子学进展*, 2020, 57(6): 061506.
- [12] Wei Y M, Quan J C, Hou Y. Aerial image location of unmanned aerial vehicle based on YOLO v2[J]. *Laser & Optoelectronics Progress*, 2017, 54(11): 111002.

- 魏湧明, 全吉成, 侯宇青阳. 基于 YOLO v2 的无人机航拍图像定位研究[J]. 激光与光电子学进展, 2017, 54(11): 111002.
- [13] Ionescu R T, Khan F S, Georgescu M I, et al. Object-centric auto-encoders and dummy anomalies for abnormal event detection in video[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE Press, 2019: 7834-7843.
- [14] Baur C, Wiestler B, Albarqouni S, et al. Deep autoencoding models for unsupervised anomaly segmentation in brain MR images[M]//Crimi A, Bakas S, Kuijf H, et al. Brainlesion: glioma, multiple sclerosis, stroke and traumatic brain injuries. Lecture notes in computer science. Cham: Springer, 2019, 11383: 161-169.
- [15] Zhou C, Paffenroth R C. Anomaly detection with robust deep autoencoders[C]//Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, August 13-17, 2017, Halifax, NS, Canada. New York: ACM, 2017: 665-674.
- [16] Akcay S, Atapour-Abarghouei A, Breckon T P. GANomaly: semi-supervised anomaly detection via adversarial training[M]//Jawahar C V, Li H D, Mori G, et al. Computer vision-ACCV 2018. Lecture notes in computer science. Cham: Springer, 2019, 11363: 622-637.
- [17] Akcay S, Atapour-Abarghouei A, Breckon T P. Skip-GANomaly: skip connected and adversarially trained encoder-decoder anomaly detection[C]//2019 International Joint Conference on Neural Networks (IJCNN), July 14-19, 2019, Budapest, Hungary. New York: IEEE Press, 2019: 1-8.
- [18] Gong D, Liu L Q, Le V, et al. Memorizing normality to detect anomaly: memory-augmented deep autoencoder for unsupervised anomaly detection [C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV), October 27-November 2, 2019, Seoul, Korea (South). New York: IEEE Press, 2019: 1705-1714.
- [19] Park H, Noh J, Ham B. Learning memory-guided normality for anomaly detection[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 13-19, 2020, Seattle, WA, USA. New York: IEEE Press, 2020: 14360-14369.
- [20] Chen Y R, Zhang H, Wang Y N, et al. MAMA net: multi-scale attention memory autoencoder network for anomaly detection[J]. IEEE Transactions on Medical Imaging, 2021, 40(3): 1032-1041.
- [21] Dai B, Guo R, Kumar S, et al. Stochastic generative hashing[C]//International Conference on Machine Learning, August 6-11, 2017, Sydney, NSW, Australia. [S.l.: s.n.], 2017: 913-922.
- [22] Bergmann P, Fauser M, Sattlegger D, et al. MVTEC AD: a comprehensive real-world dataset for unsupervised anomaly detection[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE Press, 2019: 9584-9592.
- [23] Sabokrou M, Pourreza M, Fayyaz M, et al. AVID: adversarial visual irregularity detection[M]//Jawahar C V, Li H D, Mori G, et al. Computer vision-ACCV 2018. Cham: Springer, 2019, 11366: 488-505.
- [24] Bergmann P, Löwe S, Fauser M, et al. Improving unsupervised defect segmentation by applying structural similarity to autoencoders[C]//Proceedings of the 14th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications, February 25-27, 2019. Prague, Czech Republic. SCITEPRESS-Science and Technology Publications, 2019: 372-380.
- [25] Schlegl T, Seeböck P, Waldstein S M, et al. Unsupervised anomaly detection with generative adversarial networks to guide marker discovery[M]//Niethammer M, Styner M, Aylward S, et al. Information processing in medical imaging. Lecture notes in computer science. Cham: Springer, 2017, 10265: 146-157.
- [26] Ruff L, Vandermeulen R, Goernitz N, et al. Deep one-class classification[C]//International conference on machine learning, July 10-15, 2018, Stockholm, Sweden. [S.l.: s.n.], 2018: 4393-4402.
- [27] Dehaene D, Frigo O, Combrexelle S, et al. Iterative energy-based projection on a normal data manifold for anomaly localization[C]//International Conference on Learning Representations, April 26-30, 2020, Addis Ababa, Ethiopia. [S.l.: s.n.], 2020.
- [28] Golan I, El-Yaniv R. Deep anomaly detection using geometric transformations[EB/OL]. (2018-05-28) [2021-02-05]. <https://arxiv.org/abs/1805.10917>.
- [29] Huang C Q, Ye F, Cao J K. Attribute restoration framework for anomaly detection[EB/OL]. (2019-11-25) [2021-02-05]. <https://arxiv.org/abs/1911.10676>.