计算机科学
COMPUTER SCIENCE

基于注意力的多尺度蒸馏异常检测

乔虹, 邢红杰

引用本文

乔虹, 邢红杰. 基于注意力的多尺度蒸馏异常检测[J]. 计算机科学, 2024, 51(6A): 230300223-11.

QIAO Hong, XING Hongjie. Attention-based Multi-scale Distillation Anomaly Detection[J]. Computer Science, 2024, 51(6A): 230300223-11.

计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.230300223

基于注意力的多尺度蒸馏异常检测

乔 虹 邢红杰
河北大学数学与信息科学学院河北省机器学习与计算智能重点实验室 河北 保定 071002
(qh12901@163.com)

摘 要 基于知识蒸馏的异常检测方法中,教师网络远大于学生网络,使得所得特征表示在同一位置对应图像的感受野不同。为解决此问题,可使学生网络与教师网络结构相同。然而,学生与教师网络完全相同,使得在测试阶段,对于异常样本,教师网络与学生网络特征表示差异过小而影响异常检测的性能。为解决该问题,提出了基于高效通道注意力模块的多尺度知识蒸馏异常检测方法(ECA Based Multi-Scale Knowledge Distillation Anomaly Detection,ECA-MSKDAD),并结合数据增强操作提出了相对距离损失函数。使用经过预训练的网络作为教师网络,同时使用与教师网络结构相同的网络作为学生网络。在训练阶段,对训练样本采取数据增强操作以扩充训练集的规模,并在学生网络中引入高效通道注意力(Efficient Channel Attention,ECA)模块,以增加教师网络和学生网络之间的差异,增大异常数据的重构误差,进而提高模型的检测性能。此外,利用相对距离损失函数,将数据间关系从教师网络传递到学生网络,对学生网络的网络参数进行优化。在MVTec AD进行实验,与9种相关方法比较,所提方法在异常检测与异常定位上均取得更优的性能。

关键词:深度学习;异常检测;异常定位;知识蒸馏;注意力机制
中图分类号 TP391.4

Attention-based Multi-scale Distillation Anomaly Detection

QIAO Hong and XING Hongjie
Hebei Key Laboratory of Machine Learning and Computational Intelligence,College of Mathematics and Information Science,Hebei University,Baoding,Hebei 071002,China

Abstract In the anomaly detection method based on knowledge distillation,the teacher network is much larger than the student network,so that the obtained feature representation has different visual fields corresponding to the image at the same position. In order to solve this problem,the structure of student network and teacher network can be the same. However,However,in the testing phase,the same student network and teacher network will lead to too small difference in their feature representation,which will affect the performance of anomaly detection. In order to solve this problem,ECA based multi-scale knowledge distillation anomaly detection(ECA-MSKDAD) is proposed,and a relative distance loss function is proposed based on data enhancement operation. The pre-trained network is used as the teacher network,and the network with the same network structure as the teacher network is used as the student network. In the training stage,the data enhancement operation is adopted for the training samples to expand the scale of the training set,and the efficient channel attention(ECA) module is introduced into the student network to increase the difference between the teacher network and the student network,increase the reconstruction error of the abnormal data and improve the detection performance of the model. In addition,the relative distance loss function is used to transfer the relationship between data from the teacher network to the student network,and the network parameters of the student network are optimized. Experiments on MVTec AD show that compared with nine related methods,the proposed method achieves better performance in anomaly detection and anomaly localization.

Keywords Deep learning,Anomaly detection,Abnormal location,Knowledge distillation,Attention mechanism

1 引言

在异常检测任务中,模型仅使用正常数据按无监督学习方式对单类分类器进行训练,并利用所得单类分类器将待测样本分类为正常数据或异常数据[1-2]。近年来,异常检测方法在工业图像故障检测[3]、医学图像分类[4]、网络安全检测[5]等应用领域取得了成功应用。值得一提的是,在工业图像检测应用中,异常定位是一项极具挑战性的任务,它通过为图像中每个像素计算相应的异常得分,取得精确且具有可解释性的结果。然而,在实际应用中,异常数据往往具有多样性和稀有性的特点,如工业检测中存在多种不常见的缺陷,医学图像中的视网膜及肿瘤图像较为罕见等。此外,由于异常检测任务

基金项目:国家自然科学基金(61672205);河北省自然科学基金(F2017201020);河北大学高层次人才科研启动项目(521100222002)
This work was supported by the National Natural Science Foundation of China(61672205),Natural Science Foundation of Hebei Province(F2017201020) and High-Level Talents Research Start-Up Project of Hebei University(521100222002).
通信作者:邢红杰(hjxing@hbu.edu.cn)

中的训练集仅由正常数据构成，异常数据缺失，因此常用的有监督分类方法，如多细胞多任务卷积神经网络[6]（M^2CNN）和Faster R-CNN[7]，不能用于处理异常检测任务。

现有的无监督异常检测方法往往分为两类：浅层异常检测方法和基于深度学习的异常检测方法。单类支持向量机[8]（One-Class Support Vector Machine，OCSVM）和支持向量数据描述[9]（Support Vector Data Description，SVDD）是两种常用的浅层异常检测方法。OCSVM通过最大化正常数据在特征空间中的像与原点之间的间隔构造分离超平面，在测试阶段根据待测样本的像与超平面之间的距离，将待测样本分类为正常数据或异常数据。SVDD在特征空间中求取包围正常数据的像的最小超球，在测试阶段根据待测样本与超球中心之间的距离，将待测样本分类为正常数据或异常数据。

在基于深度学习的异常检测方法中，基于重构的方法最为常用，如基于自编码器（Autoencoder，AE）[10]和基于生成式对抗网络（Generative Adversarial Nets，GAN）的异常检测方法。前者通过最小化输入样本和重构样本之间的重构误差对模型进行优化，在测试阶段，若输入样本和重构样本之间的重构误差大于预定义的阈值，则将其判定为异常数据。GANomaly[11]是近年提出的一种基于GAN的异常检测方法，在训练阶段，利用训练样本的重构损失对其生成器、编码器和判别网络进行训练，在测试阶段，根据待测样本的重构误差和预定义的阈值，将其分类为正常数据或异常数据。然而，上述基于重构的方法不仅能够有效地重构正常数据，且异常数据及其重构样本之间的重构误差往往也很小，从而将异常数据错误地判定为正常数据，进而导致模型检测性能不够理想[12]。

近年来，由Hinton等[13]提出的知识蒸馏取得了广泛关注，通常在大规模数据集上对其教师网络进行预训练，将网络输出用作知识，用于指导已完成初始化的学生网络的训练。此外，所提知识蒸馏的教师网络在计算网络输出时引入了温度参数，将硬标签转化为软标签，从而使学生网络获得更多的类别信息，因此，知识蒸馏适用于处理多类分类问题。为了处理异常检测任务，Bergmann等[14]提出了一种基于判别潜在表示的教师-学生框架，所提框架的教师网络在ImageNet数据集上进行预训练，其网络权重在预训练完成后保持不变，在训练阶段，通过对教师网络在正常数据上的特征表示进行回归，完成学生网络集成模型的训练。在测试阶段，若教师网络和学生网络在待测样本上所产生的特征表示差异较大，且集成中多个学生网络的特征表示之间的差异也较大，则将该待测样本判定为异常数据。为了充分利用网络中间层的特征表示信息，Salehi等[15]提出了一种用于异常检测的多分辨率知识蒸馏方法，所提方法在训练阶段使用教师网络中多个中间层的特征表示指导学生网络的训练，并在测试阶段利用两种网络中间层特征表示的差异进行异常检测。

然而，上述基于知识蒸馏的异常检测方法中，若教师网络与学生网络结构不同，则教师网络与学生网络所得特征表示对应输入图像的感受野不同，如图1所示。当教师网络与学生网络结构相同时，由于深度神经网络鲁棒性较强，使得对于异常样本，学生网络输出特征表示与教师网络差异较小，影响异常判断的效果。因此，引入了高效通道注意力[16]（Efficient Channel Attention，ECA）模块。ECA不仅能够增加教师网络与学生网络之间的差异，还能获取通道注意力，提高学生网络提取特征表示的能力；此外，结合数据增强，提出了相对距离损失函数，以有效提高学生网络提取特征表示的能力；提出了基于ECA的多尺度知识蒸馏异常检测（ECA Based Multi-Scale Knowledge Distillation Anomaly Detection，ECA-MSKDAD）方法。主要贡献如下：

（1）在学生网络中增加ECA模块，能够增加学生网络与教师网络之间的差异。此外，ECA模块能够提高学生网络提取特征表示的能力，从而提高异常定位的效果。

（2）在知识蒸馏过程中，充分利用多种尺度的特征表示，即教师网络利用多个中间层的所得特征表示指导学生网络的训练，并在测试阶段利用教师网络和学生网络中多个中间层的所得特征表示计算异常图，能够更精确地对异常区域进行定位。

（3）利用多种数据增强方法（包括随机灰度、颜色失真、高斯滤波）对训练样本进行扩充，有效解决学生网络在训练过程中的过拟合问题；进一步提出了相对距离损失函数，同时利用原始样本和增强样本计算两幅图像在教师网络之间的距离和在学生网络之间的距离，使得原始样本所得特征表示之间的相对距离与原始样本之间的相对距离保持一致，即原始样本之间的相对距离较小，则它们相对应的特征表示之间的相对距离亦较小。

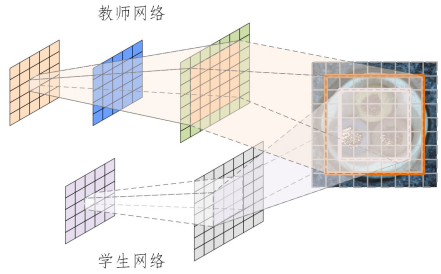


图1 教师网络的网络规模大于学生网络的网络规模时，相同特征表示所对应的不同感受野

Fig. 1 Different receptive fields corresponding to the same feature representation when the size of teacher network is larger than that of student network

2 相关工作

如上所述，OCSVM[8]和SVDD[9]是两种常用的浅层异常检测方法。除了这两种核方法，Hoffmann[17]将核主成分分析（Kernel Principal Component Analysis，KPCA）用于异常检测，首先将训练样本从输入空间映射到高维的特征空间，然后在特征空间中进行主成分分析，并将异常度量表示为特征空间中的重构误差。实验结果表明，KPCA取得了优于OCSVM的异常检测性能。Ting等[18]提出了孤立分布核（Isolation Distributional Kernel，IDK），用于度量两个分布之间的相似性。此外，将核均值嵌入中的点核替换为分布核，有效克服了点核不适于处理高维数据且与数据无关的缺点。从理论和实验两个角度证明了基于IDK的核均值嵌入在应用于异常检测时优于OCSVM和其他基于高斯核的核均值嵌入。Qu等[19]提出了基于高斯混合模型的高光谱图像异常检测方法，通过分析相邻波段之间的相关性，将高光谱图像划分为一些波段子集，然后对所得波段子集进行融合以实现维数约减，最后利用所提出的基于高斯混合模型的异常提取方法提取每个波段的

异常像素，并利用基于高斯混合模型的加权方法构造异常映射图。Li 等[20]提出了用于处理机器监测数据（Machine Monitoring Data，MMD）的相似性度量孤立森林。受益于滑动窗口的优点，所提方法在处理 MMD 时有效提高了传统孤立森林的适应能力和稳定性。此外，在异常识别阶段对疑似异常片段的相对相似性进行度量，有效提高了孤立森林的鲁棒性。Peng 等[21]在极限学习机（Extreme Learning Machine，ELM）和互信息（Mutual Information，MI）的基础之上，提出了多变量 ELM-MI 框架，并与动态核选取方法相结合。所提方法能够实现多种异常的无监督在线检测，并有效减少了计算消耗。Pang 等[22]提出了一种结合矢量量化和 OCSVM 的混合算法（VQ-OCSVM）。首先利用矢量量化提取正常数据的分布信息，利用所得结果构造显式映射函数，将数据映射到高维的特征空间；然后在特征空间中利用 OCSVM 构造单类分类器。所提方法有效解决了 OCSVM 的核参数选取问题。

虽然浅层异常检测方法在处理小规模数据集上取得了成功应用，但是在处理大规模数据集时，往往会遇到计算可扩展性差和维数灾难等问题，导致检测性能较差[23]。相较而言，基于深度学习的异常检测方法更适用于处理大规模数据集。Pang 等[24]从模型建立角度将深度异常检测划分为 3 个概念层次：用于特征提取的深度学习，学习正常数据的特征表示，端到端异常得分学习。

（1）用于特征提取的深度学习方法首先利用深度神经网络将复杂的高维数据映射到低维空间中，然后使用低维空间中所得的特征表示进行异常检测。Koppikar 等[25]提出了基于 VGG-16 和 Inception V3 的视频异常检测框架，分别利用 VGG-16 和 Inception V3 对数据进行降维，并根据降维结果对数据进行二元分类。Liznerski 等[26]提出了全卷积数据描述（Fully Convolutional Data Description，FCDD），利用全卷积网络和超球面分类器进行异常检测。全卷积网络所获得的特征表示能够保留空间信息，并且可以用作下采样异常热图；此外，FCDD 通过对低分辨率热图进行上采样，能够取得全分辨率热图。FCDD 取得了较优的检测性能，同时为常用的异常检测图像数据集提供了合理的解释。

（2）学习正常数据的特征表示的方法将特征学习与异常得分相结合，并利用损失函数计算异常得分，如自编码器（AE）、生成式对抗网络（GAN）等。Ulger 等[27]提出了基于 β-变分自编码器（β-Variational Autoencoder，β-VAE）的异常检测方法，其利用超参数 β 平衡损失函数中的重构精度和解纠缠质量，使得模型能够在保证一定重构精度的条件下具有较好的解纠缠能力。在测试阶段，所提方法利用重构损失和梯度损失的线性组合计算异常得分。Chen 等[28]提出了基于 GAN 的双自编码器异常检测模型（DAGAN），DAGAN 利用编码器-解码器-编码器构建双 GAN 模型，同时学习训练样本的潜在特征分布和边缘分布。在测试阶段，异常得分由两部分组成，分别计算两个 GAN 模型测试样本特征表示的重构误差，并对重构误差归一化，两个归一化的结果加权求和作为异常得分，以此进行异常检测。Cheng 等[29]提出了用于雷达信号时间序列的异常检测方法（ResNet-AE）。模型以自编码器为主要网络框架，其中编码器和解码器按残差结构进行堆叠，并且残差结构包含池化层、长短期记忆网络（Long Short-Term Memory，LSTM）和 ReLU 激活层。ResNet-AE 利用残差结构缓解深层网络中的梯度消失问题，且 LSTM 能够学习数据的时间相关性。在测试阶段，将待测样本的重构误差用作异常得分。

（3）端到端异常得分学习方法利用新型损失函数计算异常得分并进行异常判断，或者训练深度神经网络并进行二元分类。Li 等[30]提出了多序列学习（Multi-Sequence Learning，MSL）方法和基于 Hinge 的 MSL 排序损失。MSL 模型利用基于 Transformer 的神经网络获取视频片段的特征表示，利用所得特征表示计算异常得分，采取自训练策略对异常得分进行细化，再进一步完成异常检测。Yan 等[31]提出了混合鲁棒卷积自编码器（Hybrid Robust Convolutional Autoencoder，HRCAE），并将其应用于噪声条件下机床传感器信号的无监督异常检测。HRCAE 对平行卷积分布拟合（Parallel Convolutional Distribution Fitting，PCDF）模块进行并行训练，同时融合多个传感器的有效信息，有效提高了模型的抗噪声能力。此外，HRCAE 将所提出的融合方向距离（Fused Directional Distance，FDD）用作损失函数，综合考虑不同传感器之间的距离和角度差异，有效抑制了噪声的影响，从而提高了模型的鲁棒性。

当应用于图像异常检测时，上述深度异常检测方法仅能得到二元检测结果，无法对异常信息的位置进行标注，不具有可解释性[26]。异常定位（Anomaly Localization，AL）能够以热图形式展示像素级定位结果[32]，热图中颜色越深，异常的概率越大。

3　预备知识

本节简要介绍多尺度特征知识蒸馏、高效通道注意力和数据增强。

3.1　多尺度特征知识蒸馏

最近，Wang 等[33]提出了用于异常检测的师生金字塔特征匹配（Student-Teacher Feature Pyramid Matching Anomaly Detection，STFPM-AD）模型，其网络结构如图 2 所示。

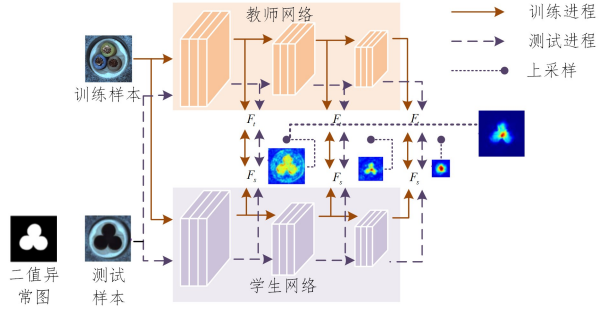


图 2　STFPM-AD 的网络结构

Fig. 2　Model structure of STFPM-AD

STFPM-AD 由教师网络和学生网络组成。教师网络采用在 ImageNet 数据集上完成预训练的 ResNet-18 网络，剔除网络中的全局平均池化层和全连接层，仅保留网络其余部分用于特征提取。学生网络采用与教师网络相同的网络结构。教师网络中具有不同大小的网络块，因而能获得不同尺度的特征表示。浅层网络得到的特征表示包含低级语义信息，深

层网络得到的特征表示具有高级语义信息。教师网络能够利用不同尺度的特征表示作为知识训练学生网络，即学生网络通过模仿教师网络输出的特征表示完成模型训练。

在 STFPM-AD 的训练过程中，将训练样本 $\boldsymbol{x}_n$ 同时输入教师网络和学生网络，利用 ℓ_2 距离计算教师网络和学生网络输出的特征表示之间的差异，并将所得差异用作损失函数，通过最小化损失函数对学生网络中的网络参数进行优化。损失函数可表示为：

$$L_{ij}^{(l)}(\boldsymbol{x}_n)=\| F_{ij}^{T(l)}(\boldsymbol{x}_n)-F_{ij}^{S(l)}(\boldsymbol{x}_n)\|_2^2 \tag{1}$$

其中，$F_{ij}^{T(l)}(\boldsymbol{x}_n)$ 表示教师网络对于样本 $\boldsymbol{x}_n$ 在第 l 层所得特征表示的 (i,j) 位置上的值，$F_{ij}^{S(l)}(\boldsymbol{x}_n)$ 表示学生网络对于样本 $\boldsymbol{x}_n$ 在第 l 层所得特征表示的 (i,j) 位置上的值。通过计算教师网络和学生网络所得特征表示在不同位置上的差异，并将所得差异加权求和，得到学生网络对于输入样本 $\boldsymbol{x}_n$ 的总损失。在训练阶段，固定教师网络的网络参数，通过最小化总损失对学生网络的网络参数进行更新。总损失公式表示如下：

$$L(\boldsymbol{x}_n)=\sum_{l=1}^{L}\alpha_l\left(\frac{1}{w_l h_l}\sum_{i=1}^{w_l}\sum_{j=1}^{h_l}L_{ij}^{(l)}(\boldsymbol{x}_n)\right) \tag{2}$$

其中，w_l 和 h_l 分别表示第 l 层输出的特征表示的宽和高，α_l 是第 l 层的损失对应的权重。

在测试阶段，学生网络对于正常数据得到的特征表示与教师网络相同；对于异常数据，学生网络与教师网络得到的特征表示有较大差异。据此，式(2)可用作异常得分。对于判断为异常的样本 $\boldsymbol{x}_t$，利用式(1)计算教师网络和学生网络第 l 层所得特征表示之间的差异，得到差异矩阵：

$$\boldsymbol{L}_{l_2}^{(l)}(\boldsymbol{x}_t)=(L_{ij}^{(l)}(\boldsymbol{x}_t))_{w_l\times h_l} \tag{3}$$

对 $\boldsymbol{L}_{\boldsymbol{x}_t}^{(l)}$ 进行上采样以获得异常图，并整合 L 个异常图，从而完成异常定位，表示为：

$$\Omega(\boldsymbol{L}_{\boldsymbol{x}_t})=\prod_{l=1}^{L}\Omega^{(l)}(\boldsymbol{L}_{\boldsymbol{x}_t}^{(l)}) \tag{4}$$

其中，$\Omega^{(l)}(\boldsymbol{L}_{\boldsymbol{x}_t}^{(l)})$ 表示对 $L_{\boldsymbol{x}_t}^{(l)}$ 进行上采样得到的异常图。

3.2 高效通道注意力

为了降低模型复杂性，压缩激励网络(Squeeze-and-Excitation Networks，SENet)[34]减少了通道数量。然而，该策略使得网络无法对权重向量与输入数据之间的对应关系进行重建，使得模型无法捕捉复杂的全局信息。为了解决该问题，Wang 等[16]提出了高效通道注意力(Efficient Channel Attention，ECA)。ECA 是一种跨通道的交互方法，该模块没有使用降维，而是使用了一维卷积进行通道之间的信息交互，其模型结构如图 3 所示。ECA 仅考虑每个通道与其 k 个近邻之间的直接信息交互，以控制模型的复杂性。ECA 可公式化为：

$$\boldsymbol{s}=F_{\text{eca}}(\boldsymbol{X},\theta)=\sigma(Conv1D(GAP(\boldsymbol{X}))) \tag{5}$$

$$\boldsymbol{Y}=\boldsymbol{s}\boldsymbol{X} \tag{6}$$

其中，$Conv1D(\cdot)$ 表示卷积核大小为 k 的一维卷积，以模拟局部跨通道交互；$GAP(\cdot)$ 表示全局平均池化层；σ 表示 Sigmoid 激活函数。参数 k 的大小决定交互的范围。在 ECA 模块中，k 的通道维度 C 无需通过交叉验证手动调整，而是能够自适应地确定，即：

$$k=\psi(C)=\left|\frac{\log_2(C)}{\gamma}+\frac{b}{\gamma}\right|_{\text{odd}} \tag{7}$$

其中，$\gamma=2$ 和 $b=1$ 为超参数，$|\cdot|_{\text{odd}}$ 表示距离式中数字最近的奇数。

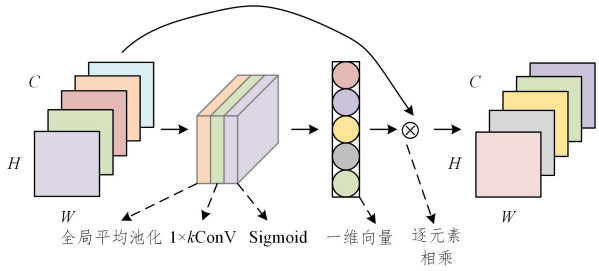


图 3 ECA 模型结构程

Fig. 3 Model structure of ECA

3.3 数据增强

数据增强是指通过对现有数据进行一系列变换操作，以生成更多的训练样本来扩充数据集的技术[35]。这些变换操作可以是随机的，也可以是基于先验知识提前设计的。通过数据增强可以提高模型的泛化能力和鲁棒性，减小模型对数据集中噪声和局部特征的依赖性，防止模型出现过拟合现象。同时，数据增强还可以增加数据集的多样性，提高模型的适应性和识别能力。常见的数据增强操作包括旋转、翻转、缩放、裁剪、平移、变形、噪声、颜色变换等。这些操作可以单独或组合使用，根据具体应用场景和需求进行选择和调整。

数据增强可以根据其实现方式和应用场景分为以下 4 类[36]。

(1)几何变换的方法：包括旋转、平移、缩放、翻转、裁剪、变形等，这些方法主要用于图像处理任务，可以改变图像的形态、大小、位置等。

(2)颜色变换的方法：包括颜色抖动、色彩偏移、对比度调整、亮度调整等，这些方法主要用于图像分类、检测等任务中，可以增加数据集的多样性和鲁棒性。

(3)图像混合的方法：包括随机遮挡、Mixup、CutMix 等，这些方法主要用于增加数据集的多样性和鲁棒性，同时可以减少过拟合现象的发生。

(4)深度学习的方法：包括利用神经网络迁移图像风格、对抗训练、生成对抗网络等，这些方法主要用于利用数据集原始特征信息，增加更贴近原数据集风格的图像。

4 基于注意力的多尺度蒸馏异常检测

本节将从模型结构和损失函数两个角度对所提方法进行详细描述。

4.1 模型结构

ECA-MSKDAD 采用教师-学生的知识蒸馏框架。教师网络使用在 ImageNet 数据集上经过预训练的 ResNet-18 网络，在原始网络结构的基础上去除了 ResNet-18 中最后的平均池化层和全连接层。当教师网络的网络规模远远大于学生网络的网络规模时，两种网络所得特征表示的相同元素所对应的图像视野范围相差很大，如图 1 所示，使得学生网络的训练不够充分，进而产生较差的检测性能。为了解决该问题，可采用与教师网络结构相同的网络作为学生网络。然而，当教师网络与学生网络结构完全相同时，两种网络得到的特征表示亦完全相同，因此学生网络在异常数据上会取得较小的重构误差，产生误判，进而取得较差的检测性能。因此本文采用与教师网络结构相同的学生网络，并在此基础上对学生网络加以改造，即在学生网络中的一部分批归一化层后添加 ECA

模块，以增加教师网络和学生网络之间的差异，增大异常数据的重构误差，进而提高模型的检测性能。

在训练阶段，ECA-MSKDAD 的模型结构如图 4 所示。教师网络参数保持不变，学生网络参数进行随机初始化。训练时给定仅包含正常样本的样本数为 N 的数据集 $D=\{\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_N\}$，对于任意的训练样本 $\boldsymbol{x}_n$，利用数据增强方法生成对应的增强样本 $\boldsymbol{x}_n^{(a)}$，以此增加训练样本数量。增强样本 $\boldsymbol{x}_n^{(a)}$ 可通过对样本 $\boldsymbol{x}_n$ 进行颜色失真、随机灰度或高斯滤波操作得到。将 $\boldsymbol{x}_n$ 分别输入教师网络和学生网络，计算其通过教师网络和学生网络所得特征表示之间的差异，得到样本 $\boldsymbol{x}_n$ 的特征损失；同理可计算增强样本 $\boldsymbol{x}_n^{(a)}$ 通过教师网络和学生网络所得特征表示之间的差异，得到增强样本 $\boldsymbol{x}_n^{(a)}$ 的特征损失，将样本 $\boldsymbol{x}_n$ 的特征损失和增强样本 $\boldsymbol{x}_n^{(a)}$ 的特征损失求和得到总的特征损失。计算教师网络中训练样本 $\boldsymbol{x}_n$ 和增强样本 $\boldsymbol{x}_n^{(a)}$ 之间的差异，同理计算学生网络中训练样本 $\boldsymbol{x}_n$ 和增强样本 $\boldsymbol{x}_n^{(a)}$ 之间的差异，并最小化两种差异间的距离。将上述总的特征损失与所求的距离求和，通过最小化求和结果优化学生网络的网络参数。

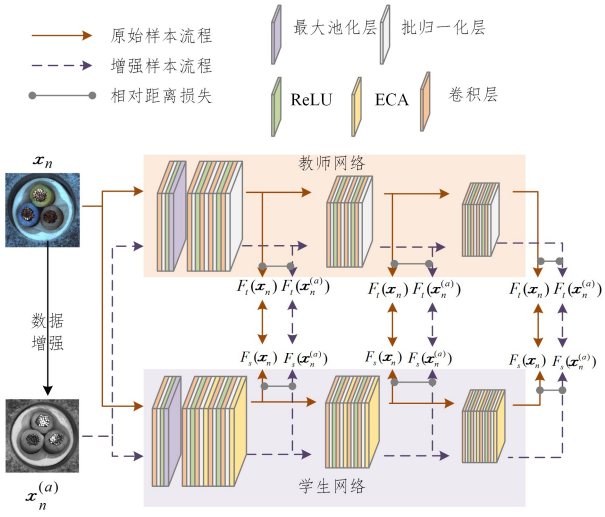


图 4　ECA-MSKDAD 在训练阶段的模型结构

Fig. 4　Model structure of ECA-MSKDAD in training phase

在测试阶段，ECA-MSKDAD 的模型结构如图 5 所示。将测试样本 $\boldsymbol{x}_t$ 输入教师网络和学生网络，利用教师网络和学生网络所得特征表示之间的差异进行异常判断。对于判断为异常的样本，利用式(12)计算第 l 层的异常图，将不同大小的异常图进行上采样，使其与输入图像大小相等，然后利用式(4)计算最终的异常定位图。

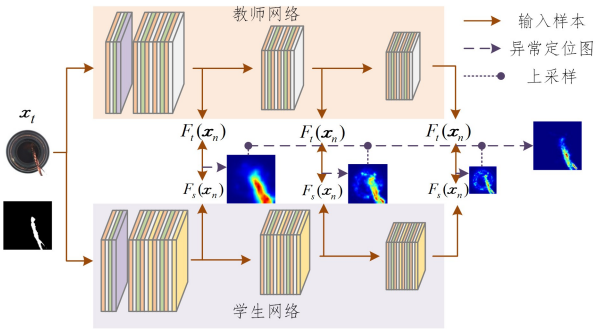


图 5　ECA-MSKDAD 在测试阶段的模型结构

Fig. 5　Model structure of ECA-MSKDAD in test phase

4.2　损失函数

4.2.1　特征图损失函数

ECA-MSKDAD 在训练阶段的目标是获得能够较好模仿教师网络输出结果的学生网络。给定仅包含正常样本的样本数为 N 的数据集 $D=\{\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_N\}$，任意样本 $\boldsymbol{x}_n\in\mathbb{R}^{h\times w\times c}$，其中 h，w 和 c 分别为图像的高度、宽度和通道数。训练过程中，学生网络模仿教师网络输出的不同尺度的特征表示，对于任意输入图像 $\boldsymbol{x}_n$，教师网络在第 l 层输出的特征表示为 $F^{T(l)}(\boldsymbol{x}_n)\in\mathbb{R}^{h_l\times w_l\times c_l}$；同理，学生网络在第 l 层输出的特征表示为 $F^{S(l)}(\boldsymbol{x}_n)\in\mathbb{R}^{h_l\times w_l\times c_l}$，其中 h_l，w_l 和 c_l 分别表示第 l 层输出特征表示的高度、宽度和通道数。对于特征图中每个位置 (i,j)，学生网络输出的特征表示 $F_{ij}^{S(l)}(\boldsymbol{x}_n)$ 模仿教师网络输出的特征表示 $F_{ij}^{T(l)}(\boldsymbol{x}_n)$，其中 $F_{ij}^{T(l)}(\boldsymbol{x}_n)\in\mathbb{R}^{c_l}$ 和 $F_{ij}^{S(l)}(\boldsymbol{x}_n)\in\mathbb{R}^{c_l}$ 分别表示教师网络和学生网络在第 l 层输出的特征表示在位置 (i,j) 处的特征向量。与大多知识蒸馏方法类似，对于样本 $\boldsymbol{x}_n$，教师网络和学生网络在第 l 层输出的特征表示在位置 (i,j) 经过归一化后所得向量间的差异为：

$$L_{l_2,ij}^{(l)}(\boldsymbol{x}_n)=\frac{1}{2}\left\|\hat{F}_{ij}^{T(l)}(\boldsymbol{x}_n)-\hat{F}_{ij}^{S(l)}(\boldsymbol{x}_n)\right\|_2^2 \tag{8}$$

其中，$\hat{F}_{ij}^{T(l)}(\boldsymbol{x}_n)=\dfrac{F_{ij}^{T(l)}(\boldsymbol{x}_n)}{\left\|F_{ij}^{T(l)}(\boldsymbol{x}_n)\right\|_2}$，$\hat{F}_{ij}^{S(l)}(\boldsymbol{x}_n)=\dfrac{F_{ij}^{S(l)}(\boldsymbol{x}_n)}{\left\|F_{ij}^{S(l)}(\boldsymbol{x}_n)\right\|_2}$。进一步由式(3)可得差异矩阵，记为 $\boldsymbol{L}_{l_2}^{(l)}(\boldsymbol{x}_n)$。

除了使用 ℓ_2 距离，为了充分利用向量间的角度差异，也采用余弦相似性度量计算 $\hat{F}_{ij}^{T(l)}(\boldsymbol{x}_n)$ 和 $\hat{F}_{ij}^{S(l)}(\boldsymbol{x}_n)$ 之间的差异，即：

$$L_{\cos,ij}^{(l)}(\boldsymbol{x}_n)=1-\frac{\hat{F}_{ij}^{T(l)}(\boldsymbol{x}_n)\cdot\hat{F}_{ij}^{S(l)}(\boldsymbol{x}_n)}{\left\|\hat{F}_{ij}^{T(l)}(\boldsymbol{x}_n)\right\|_2\left\|\hat{F}_{ij}^{S(l)}(\boldsymbol{x}_n)\right\|_2} \tag{9}$$

由式(9)进一步可得余弦相似度矩阵，即：

$$\boldsymbol{L}_{\cos}^{(l)}(\boldsymbol{x}_n)=(L_{\cos,ij}^{(l)}(\boldsymbol{x}_n))_{w_l\times h_l} \tag{10}$$

因此对于样本 $\boldsymbol{x}_n$，学生网络在第 l 层获得的特征表示在位置 (i,j) 处的特征差异矩阵表示为 ℓ_2 距离差异矩阵与余弦相似度矩阵的和，表示为：

$$\boldsymbol{L}^{(l)}(\boldsymbol{x}_n)=\boldsymbol{L}_{l_2}^{(l)}(\boldsymbol{x}_n)+\lambda\boldsymbol{L}_{\cos}^{(l)}(\boldsymbol{x}_n) \tag{11}$$

其中，λ 是折中参数，用于平衡 ℓ_2 距离和余弦相似度。

对于样本 $\boldsymbol{x}_n$，学生网络在第 l 层特征表示的特征损失定义为所有位置上损失的平均值，即：

$$\boldsymbol{L}^{(l)}(\boldsymbol{x}_n)=\frac{1}{w_l h_l}\sum_{i=1}^{h_l}\sum_{j=1}^{w_l}\boldsymbol{L}^{(l)}(\boldsymbol{x}_n) \tag{12}$$

同理，对于增强样本 $\boldsymbol{x}_n^{(a)}$，学生网络在第 l 层特征表示的特征损失表示为：

$$L^{(l)}(\boldsymbol{x}_n^{(a)})=\frac{1}{w_l h_l}\sum_{i=1}^{h_l}\sum_{j=1}^{w_l}\boldsymbol{L}^{(l)}(\boldsymbol{x}_n^{(a)}) \tag{13}$$

对于样本 $\boldsymbol{x}_n$ 和其对应的增强样本 $\boldsymbol{x}_n^{(a)}$，将不同层的特征损失加权求和，可得总的特征损失为：

$$L_{\text{feature}}=\sum_{l=1}^{L}\alpha_l(L^{(l)}(\boldsymbol{x}_n)+L^{(l)}(\boldsymbol{x}_n^{(a)})) \tag{14}$$

其中，α_l 表示第 l 层特征损失的权重，在实验中设置为 $\alpha_l=1$ $(l=1,\cdots,L)$。

4.2.2　相对距离损失函数

现有的基于蒸馏的异常检测方法仅利用特征表示之间的差异对学生网络的网络参数进行优化，因此只能将基于特征表示的知识传递给学生网络[37]，而并未考虑数据之间的关系。为了解决该问题，提出了相对距离损失函数，其直观解释

如图 6 所示。通过最小化原样本 $\boldsymbol{x}_n$ 和增强样本 $\boldsymbol{x}_n^{(a)}$ 经教师网络所得特征表示的差异与经学生网络所得特征表示的差异之间的距离，能够将原样本 $\boldsymbol{x}_n$ 和增强样本 $\boldsymbol{x}_n^{(a)}$ 之间的关系从教师网络传递到学生网络。值得一提的是，增强样本 $\boldsymbol{x}_n^{(a)}$ 是由原样本 $\boldsymbol{x}_n$ 经过概率均为 0.8 的颜色失真、随机灰度变化和高斯滤波处理后得到的增强图像。

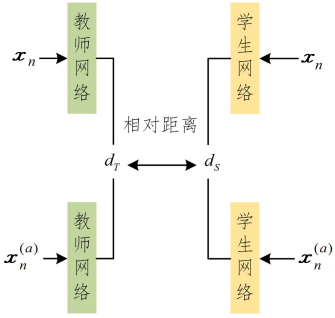


图 6 相对距离损失

Fig. 6 Relative distance loss

在相对距离损失函数中，将原样本 $\boldsymbol{x}_n$ 与增强样本 $\boldsymbol{x}_n^{(a)}$ 通过教师网络第 l 层所得特征表示之间的距离记为 $d_T^{(l)}$，即：

$$d_T^{(l)} = \| F^{T(l)}(\boldsymbol{x}_n) - F^{T(l)}(\boldsymbol{x}_n^{(a)}) \|_2^2 \tag{15}$$

将原样本 $\boldsymbol{x}_n$ 与增强样本 $\boldsymbol{x}_n^{(a)}$ 通过学生网络第 l 层所得特征表示之间的距离记为 $d_S^{(l)}$，即：

$$d_S^{(l)} = \| F^{S(l)}(\boldsymbol{x}_n) - F^{S(l)}(\boldsymbol{x}_n^{(a)}) \|_2^2 \tag{16}$$

通过最小化 $d_T^{(l)}$ 与 $d_S^{(l)}$ 之间的距离对学生网络的网络参数进行优化，因此相对距离损失函数可以表示为：

$$L_{\text{rel}} = \sum_{l=1}^{L} \| d_S^{(l)} - d_T^{(l)} \|_2^2 \tag{17}$$

综合考虑特征损失(式(14))和相对距离损失(式(17))，ECA-MSKDAD 的损失函数表示为：

$$L = L_{\text{feature}} + L_{\text{rel}} \tag{18}$$

5 实验及结果

为验证所提 ECA-MSKDAD 的有效性，将其与 8 种相关方法在 MVTec AD 数据集[10]上进行了实验比较并做了消融研究。

5.1 数据集及参数设置

实验中使用 MVTecAD 数据集，该数据集共包含 5 354 张高分辨率彩色图像，其中训练集包括 3 629 张图像，测试集包括 1 725 张图像。数据集中图像分为 15 个类，其中 5 类为纹理图像，10 类为物体图像，并且训练集中只包含正常样本，测试集中包含正常样本和 73 种不同的异常样本。此外，对每一个异常样本，数据集中都给出了像素级的二值异常图，可对异常定位任务中的模型性能进行衡量。MVTec AD 数据集的相关信息如表 1 所列。

表 1 MVTec AD 数据集的相关信息

Table 1 Information of MVTec AD

	类别	训练样本个数	测试样本个数(正常)	测试样本个数(异常)	异常种类	异常区域面积	图像边长
纹理	Carpet	280	28	89	5	97	1 024
	Grid	264	21	57	5	170	1 024
	Leather	245	32	92	5	99	1 024
	Tile	230	33	84	5	86	840
	Wood	247	19	60	5	168	1 024
物体	Bottle	209	20	63	3	68	900
	Cable	224	58	92	8	151	1 024
	Capsule	219	23	109	5	114	1 000
	Hazelnut	391	40	70	4	136	1 024
	Metal nut	220	22	93	4	132	700
	Pill	267	26	141	7	245	800
	Screw	320	41	119	5	135	1 024
	Toothbrush	60	12	30	1	66	1 024
	Transistor	213	60	40	4	44	1 024
	Zipper	240	32	119	7	177	1 024

MVTec AD 数据集的纹理和物体示例分别如图 7 和图 8 所示，最上方为类别名称，图中第一行表示正常数据，第二行表示异常数据，第三行表示图中第二行异常数据对应的二值异常图。

在测试阶段，首先判断图像是否异常，在该任务中，使用 AUCROC 评价指标作为图像级异常检测的评价标准。对于判断为异常的图像，利用 AUCROC 和 PRO 作为指标，将本文提出的方法与其他相关方法进行比较。

实验中，MVTec AD 图像大小设置为 128×128 像素，迭代 300 次，将批量大小设置为 32。学生网络采用 SGD 优化器，学习率设置为 0.4，采用 0.9 的动量更新。所提方法 ECA-MSKDAD 所使用的网络使用 PyTorch 框架搭建。

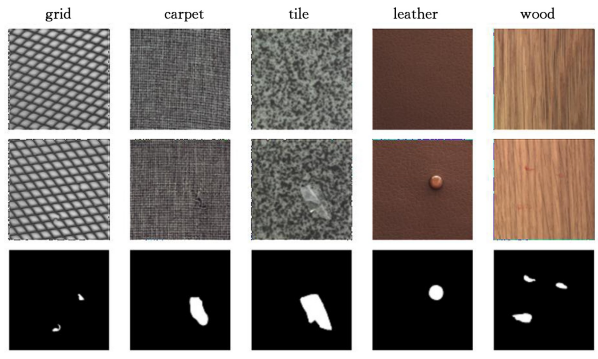


图 7 MVTec AD 中纹理对象的示例图

Fig. 7 Examples of texture in MVTec AD

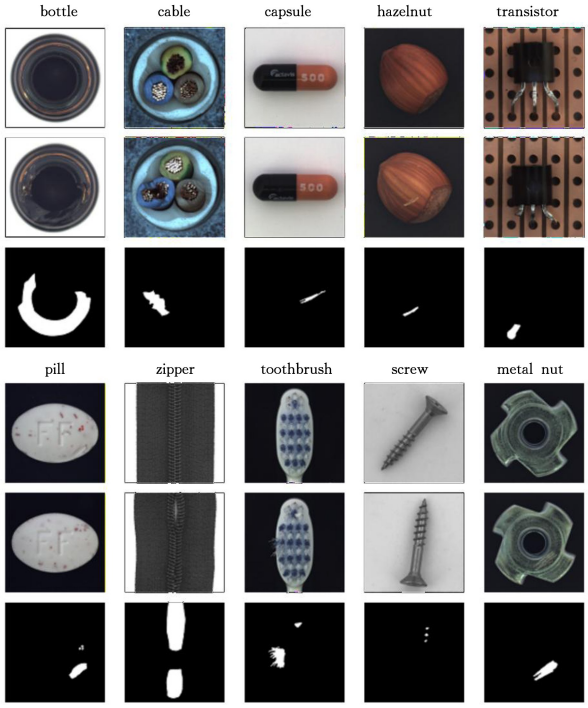


图 8　MVTec AD 中对象的示例图

Fig. 8　Examples of objects in MVTec AD

5.2　数据集及评价指标

为了验证本文所提 ECA-MSKDAD 的性能，将它与以下相关方法进行比较。

Geom[38]：该模型对训练集中的样本采取一组数据增强操作，并为采取不同增强的样本生成对应标签，以此创建自标记的多类数据集。在训练过程中，通过网络预测输入样本所属的自标签类别，增强网络模拟样本分布的能力。

GANomaly[11]：该模型使用编码器-解码器-编码器作为生成器，同时使用对抗训练的思想，通过交替优化生成器和判别器，提高网络训练速度，并根据特征表示和输入样本的重构损失优化网络参数。

ITAE[39]：该模型通过对原始样本采用数据增强操作（如灰度变化、随机旋转等），将数据增强后的图像输入自编码器，通过最小化原始样本和自编码器输出结果之间的重构误差优化自编码器。该模型提高了自编码器对语义信息提取的能力。

SSIM-AE[12]：该模型应用了基于结构相似性的感知损失函数，该损失函数利用图像局部区域之间的相互依赖性，同时考虑亮度、对比度和结构信息，代替比较单个像素值，提高了模型在异常定位任务的效果。

AnoGAN[40]：该模型是首个将 GAN 用于异常检测的方法，训练期间仅利用正常数据，促使 DCGAN 学习正常样本特征分布，并利用判别器进行判断。

CNN-Dict[41]：该模型利用从训练集图像获取的子图构成字典，在测试阶段，计算测试图像的子图与训练期间构成字典的子图的相似性，相似性越低，异常程度越高。这种利用子图的方法能够同时完成异常检测和定位。

CutPaste[42]：该模型使用两阶段训练框架构建异常检测器。首先利用数据增强操作，裁剪图像块并将图像块粘贴在原图上，利用原图像和增强图像训练网络进行二值分类，在第二阶段利用第一阶段训练的神经网络判断图像正常或异常，利用 GradCAM 定位缺陷，并利用图像块提取特征表示以生成异常图。

Patch SVDD[43]：该模型将深度支持向量数据描述扩展为基于图像块的自监督学习方法，能够在图像级异常检测的基础上完成异常分割。

PaDiM[44]：该模型在训练阶段将训练样本的 patch 块输入经过预训练的卷积神经网络，将网络不同层输出的特征表示进行连接得到嵌入特征向量，利用多元高斯分布拟合嵌入特征向量分布。在测试阶段，利用马氏距离获得图像异常分数和异常图。

SPADE[45]：该模型在训练阶段将训练样本映射到特征空间，在测试阶段，将待测样本映射到特征空间，从特征空间中检索 K 个最近的正常样本，计算待测样本与 K 个最近正常样本的欧氏距离，根据给定阈值判断样本正常或异常。对于判断为异常的样本，利用预训练的网络进行特征匹配，完成异常定位。

STAD[14]：该模型使用在 ImageNet 数据集上经过预训练的教师网络，其网络权重在预训练完成后保持不变，在训练阶段，通过对教师网络在正常数据上的潜在表示进行回归，完成学生网络集成模型的训练。在测试阶段，若教师网络和学生网络在待测样本上所产生的潜在表示差异较大，且集成中多个学生网络的潜在表示之间的差异也较大，则将该待测样本判定为异常。

InTra[46]：该模型利用 Transformer 替换卷积神经网络，使得网络能够更好地捕捉上下文的信息，通过最小化图像块的重构误差训练网络。测试阶段，根据局部和全局的重构图像完成异常检测和异常定位。

在实验中，依次选取数据集每一类作正类。训练过程仅使用训练样本（正常样本），测试过程使用数据集中的测试样本（包含正常样本和异常样本），对于判断为异常的样本，输出其不同尺度的异常定位图，并使用 AUCROC 和 PRO 作为评价指标来评估结果。

5.3　比较分析

在图像级异常检测评估中，使用 AUCROC 作为评价指标，将所提 ECA-MSKDAD 与 Geom[38]，GANomaly[11]，ITAE[39]，CutPaste[42]，Patch SVDD[43]，PaDiM-WR50[44]，SPADE[45] 和 InTa[46] 方法进行比较，图像级异常检测的 AUCROC 测试结果如表 2 所列，其中粗体为最优结果，下划线为次优结果。

表 2　9 种不同方法在 MVTec AD 数据集上的图像级 AUCROC 测试结果

Table 2　Image-level AUCROC test results of 9 different methods in MVTec AD

Geom	GANomaly	ITAE	CutPaste	PatchSVDD	PaDiM-WR50	SPADE	InTra	Ours
0.672	0.762	0.839	0.952	0.921	<u>0.953</u>	0.855	0.950	0.973

在像素级异常检测和异常定位的评估中，分别使用 AUCROC 和 PRO 作为评价指标，将所提 ECA-MSKDAD 与 SSIM-AE[12]，AnoGAN[40]，CNN-Dict[41]，CutPaste[42]，Patch SVDD[43]，PaDiM-R18[44]，SPADE[45]，STAD[14] 和 InTa[46] 方法进行比较。

上述异常检测方法在 MVTec AD 数据集上的 AUCROC 和 PRO 测试结果如表 3、表 4、图 9 和图 10 所示，其中粗体为最优结果，下划线为次优结果。

表 3 9 种不同方法在 MVTec AD 数据集上的像素级 AUCROC 测试结果

Table 3 Pixel-level AUCROC test results of 9 different methods in MVTec AD

	Category	SSIM-AE	AnoGAN	CNN-Dict	CutPaste	Patch SVDD	PaDiM-R18	SPADE	InTra	Ours
Textures	Carpet	0.87	0.54	0.72	0.983	0.926	0.989	0.975	0.992	**0.993**
	Grid	0.94	0.58	0.59	0.975	0.962	0.949	0.937	0.988	0.990
	Leather	0.78	0.64	0.87	**0.995**	0.974	0.991	0.976	**0.995**	**0.995**
	Tile	0.59	0.50	0.93	0.905	0.914	0.912	0.874	0.944	**0.982**
	Wood	0.73	0.62	0.91	0.955	0.908	0.936	0.885	0.887	**0.965**
Objects	Bottle	0.93	0.86	0.78	0.976	0.981	0.981	0.984	0.971	**0.991**
	Cable	0.82	0.78	0.79	0.900	0.968	0.958	**0.972**	0.910	0.970
	Capsule	0.94	0.84	0.84	0.974	0.958	0.983	**0.990**	0.977	0.985
	Hazelnut	0.97	0.87	0.72	0.973	0.975	0.977	**0.991**	0.983	0.990
	Metal nut	0.89	0.76	0.82	0.931	0.980	0.967	0.981	0.933	**0.985**
	Pill	0.91	0.87	0.68	0.957	0.951	0.947	0.965	0.983	**0.986**
	Screw	0.96	0.80	0.87	0.967	0.957	0.974	0.989	**0.995**	0.984
	Toothbrush	0.92	0.93	0.90	0.981	0.981	0.987	0.979	0.989	**0.992**
	Transistor	0.90	0.86	0.66	0.930	0.970	**0.972**	0.941	0.961	0.889
	Zipper	0.88	0.78	0.76	**0.993**	0.951	0.982	0.965	0.992	0.992
	Mean	0.87	0.74	0.78	0.960	0.957	0.967	0.965	0.966	**0.980**

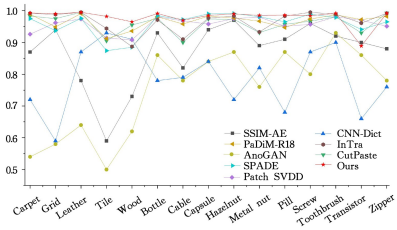


图 9 9 种不同方法在 MVTec AD 数据集上的像素级 AUCROC 测试结果

Fig. 9 Pixel-level AUCROC test results of 9 different methods in MVTec AD

由表 3 可知，ECA-MSKDAD 在 MVTec AD 数据集的纹理对象中均取得最优的 AUCROC 结果。在物体对象中，除 Transistor(晶体管)和 Screw(螺丝钉)外均取得最优和次优的结果，并且在所有类别的平均结果上取得最优的结果，显示出模型整体的优势。这是由于模型在批归一化层后增加 ECA 模块，提高了学生网络提取样本特征表示的能力，然而，由于模型所使用的 ResNet-18 网络层数较少，使得模型未能在所有物体对象中展现出最优结果。由图 9 中可直观地发现所提 ECA-MSKDAD 在各个类别上都取得了更为稳定且优异的性能。

表 4 7 种不同方法在 MVTec AD 数据集上的 PRO 测试结果

Table 4 PRO test results of 7 different methods in MVTec AD

	Category	SSIM-AE	AnoGAN	CNN-Dict	STAD	PaDiM-R18	SPADE	Ours
Textures	Carpet	0.65	0.20	0.47	0.695	**0.960**	0.947	**0.960**
	Grid	0.85	0.23	0.18	0.819	0.909	0.867	**0.967**
	Leather	0.56	0.38	0.64	0.819	0.979	0.972	**0.985**
	Tile	0.18	0.18	0.80	0.912	0.816	0.759	**0.957**
	Wood	0.61	0.39	0.62	0.74	0.918	0.939	**0.942**
Objects	Bottle	0.83	0.62	0.74	0.918	0.939	0.955	**0.957**
	Cable	0.48	0.38	0.56	0.865	0.862	0.909	**0.916**
	Capsule	0.86	0.31	0.31	0.916	0.919	0.937	**0.953**
	Hazelnut	0.92	0.70	0.84	0.937	0.914	0.954	**0.966**
	Metal nut	0.60	0.32	0.36	0.895	0.819	0.944	**0.947**
	Pill	0.83	0.78	0.46	0.935	0.906	0.946	**0.970**
	Screw	0.89	0.47	0.28	0.928	0.913	**0.960**	0.937
	Toothbrush	0.78	0.75	0.15	0.863	0.923	**0.935**	0.934
	Transistor	0.73	0.55	0.63	0.701	0.802	**0.874**	0.858
	Zipper	0.67	0.47	0.70	0.933	0.947	0.926	**0.955**
	Mean	0.69	0.44	0.52	0.857	0.901	0.917	**0.946**

由表 4 可知，ECA-MSKDAD 在 MVTec AD 数据集的纹理对象中均取得最优的 PRO 结果，在物体对象中均取得最优和次优的结果，并且在所有类别的平均结果上取得最优的结果，显示出模型整体的优势。由图 10 中可直观地发现所提 ECA-MSKDAD 在各个类别上都取得了更为稳定且优异的性能。

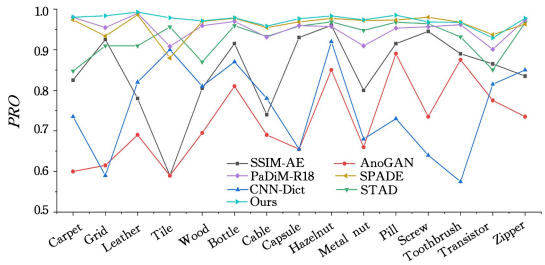


图 10　7 种不同方法在 MVTec AD 数据集上的 PRO 测试结果

Fig. 10　PRO test results of 7 different methods in MVTec AD

异常定位的可视化结果如图 11 所示，每行表示不同的异常样本，其中(a)列为异常数据，(b)列为 16×16 大小的异常定位图，(c)列为 32×32 大小的异常定位图，(d)列为 64×64 大小的异常定位图，(e)列为融合后的异常定位图，(f)列为利用异常定位图标记异常数据的结果，(g)列为二值异常图。

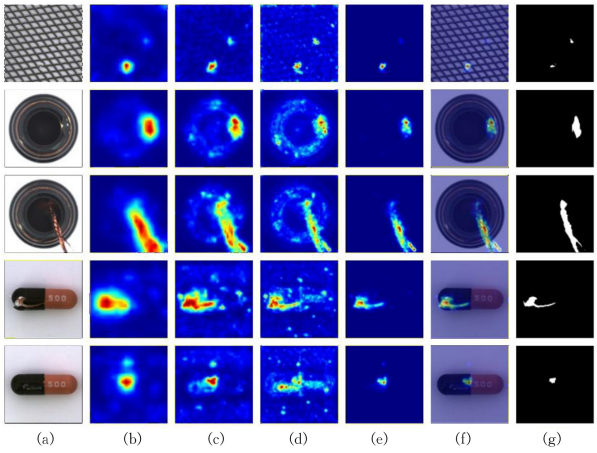


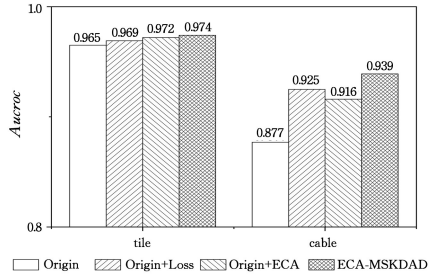
图 11　异常定位可视化结果

Fig. 11　Visualization results of anomaly location

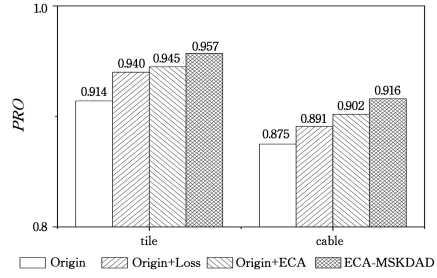
由图 11 中的结果可知，对于尺度较小的异常定位图 11(b)列，能够较为精准地捕获异常，但会忽略较小的区域，同时难以把握异常区域的边缘；尺度较大的异常定位图 11(d)列，能够更好地捕获异常区域的边缘，但会包括更多的正常区域，同时异常定位图更接近原始输入图像；尺度居中的异常定位图 11(c)列，则位于上述两种表现之间。因此将不同尺度的异常定位图融合，能够获得包含上述异常定位图优点的异常定位图。

5.4　消融研究

为了研究学生网络中部分批归一化层后增加高效通道注意力模块的有效性，以及所提出的相对距离损失函数的有效性，分别在 MVTec AD 数据集的纹理类 Tile(瓷砖)和物体类 Cable(电缆)上通过消融实验进行验证。消融实验中将未在学生网络中增加高效通道注意力模块且未使用相对距离损失函数的方法命名为 Origin，仅增加相对距离损失函数的方法命名为 Origin＋Loss，仅在学生网络部分批归一化层后增加 ECA 模块的方法命名为 Origin＋ECA，上述方法与本文所提方法 ECA-MSKDAD 的比较结果如图 12 所示。由图 12 可知，ECA-MSKDAD 在两组数据上均表现出了最佳性能。与 Origin＋Loss 相比，ECA-MSKDAD 在部分批归一化层后增加 ECA 模块，以增大学生网络与教师网络之间的差异，同时有效提高学生网络提取正常数据关键部分的能力，产生了更优的结果。与 Origin＋ECA 相比，ECA-MSKDAD 使用了相对距离损失函数，使得学生网络能够对正常数据产生更紧凑的映射，从而取得更优的检测和定位性能。



(a) AUCROC 评价指标的结果



(b) PRO 评价指标的结果

图 12　在 MVTec AD 数据集的消融实验

Fig. 12　Ablation experiment in MVTec AD

结束语　本文将基于知识蒸馏的异常检测方法与高效通道注意力模块 ECA 相结合，提出了基于 ECA 的多尺度知识蒸馏的异常检测方法，还提出了相对距离损失函数。ECA-MSKDAD 在学生网络的部分批归一化层后增加 ECA 模块，增大了教师网络与学生网络的差异。同时，结合数据增强操作提出了相对距离损失函数，通过最小化原样本和增强样本经教师网络所得特征表示的差异与经学生网络所得特征表示的差异之间的距离优化学生网络的网络参数。在 MVTec AD 数据集开展了实验，验证了所提方法的有效性。

本文所提出的方法使用了知识蒸馏的网络框架，其中教师网络使用在 ImageNet 数据集上经过预训练的网络，然而当预训练所用数据与模型实际训练所用数据有较大的差异时，可能会导致教师网络的指导能力下降。针对此问题，将尝试从以下两个方面进行探索：(1)尝试其他的知识蒸馏模型，如在线知识蒸馏模型，在训练学生网络之前，首先对教师网络进行微调，以获得更符合实际情况的教师网络；(2)探索知识蒸馏模型的不同网络框架，以获得效果更好的学生网络，提高异常检测和异常定位的性能。

参 考 文 献

[1] KHAN S S, MADDEN M G. A Survey of Recent Trends in One Class Classification[C]// Irish Conference on Artificial Intelligence and Cognitive Science. Springer, Berlin, Heidelberg, 2009: 188-197.

[2] AHMED M, MAHMOOD A N, HU J. A Survey of Network Anomaly Detection Techniques[J]. Journal of Network and Computer Applications, 2016, 60: 19-31.

[3] 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.

[4] YAO Q, XIAO L, LIU P, et al. Label-free Segmentation of Covid-19 Lesions in Lung CT[J]. IEEE Transactions on Medical Imaging, 2021, 40(10): 2808-2819.

[5] RASHID A N M B, AHMED M, SIKOS L F, et al. Anomaly Detection in Cybersecurity Datasets Via Cooperative Co-Evolution-Based Feature Selection[J]. ACM Transactions on Management Information Systems(TMIS), 2022, 13(3): 1-39.

[6] ZHOU K, GU Z, LIU W, et al. Multi-Cell Multi-Task Convolutional Neural Networks for Diabetic Retinopathy Grading[C]// 2018 40th Annual International Conference of The IEEE Engineering in Medicine and Biology Society(EMBC). IEEE, 2018: 2724-2727.

[7] DHIMAN G, JUNEJA S, VIRIYASITAVAT W, et al. A Novel Machine-Learning-Based Hybrid CNN Model for Tumor Identification in Medical Image Processing[J]. Sustainability, 2022, 14(3): 1447.

[8] SCHÖLKOPF B, WILLIAMSON R C, SMOLA A, et al. Support Vector Method for Novelty Detection[J]. Advances in Neural Information Processing Systems, 1999, 12.

[9] TAX D M J, DUIN R P W. Support Vector Data Description [J]. Machine Learning, 2004, 54(1): 45-66.

[10] 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.

[11] AKCAY S, ATAPOUR-ABARGHOUEI A, BRECKON T P. GANomaly: Semi-Supervised Anomaly Detection Via Adversarial Training[C]// Computer Vision-ACCV 2018: 14th Asian Conference on Computer Vision. Springer International Publishing, 2019: 622-637.

[12] BERGMANN P, LWE S, FAUSER M, et al. Im-proving Unsupervised Defect Segmentation by Applying Structural Similarity to Autoencoders[C]// 14th International Conference on Computer Vision Theory and Applications. 2019.

[13] HINTON G, VINYALS O, DEAN J. Distilling the Knowledge in a Neural Network[J]. Computer Science, 2015, 14(7): 38-39.

[14] 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.

[15] SALEHI M, SADJADI N, BASELIZADEH S, et al. Multiresolution Knowledge Distillation for Anomaly Detection[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021: 14902-14912.

[16] WANG Q, WU B, ZHU P, et al. ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks[C]// 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR). 2020: 11531-11539.

[17] HOFFMANN H. Kernel PCA for Novelty Detection[J]. Pattern Recognition, 2007, 40(3): 863-874.

[18] TING K M, XU B C, WASHIO T, et al. Isolation Distributional Kernel: A New Tool for Kernel Based Anomaly Detection[C]// Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. 2020: 198-206.

[19] QU J, DU Q, LI Y, et al. Anomaly Detection in Hyperspectral Imagery Based on Gaussian Mixture Model[J]. IEEE Transactions on Geoscience and Remote Sensing, 2020, 59(11): 9504-9517.

[20] LI C, GUO L, GAO H, et al. Similarity-Measured Isolation Forest: Anomaly Detection Method for Machine Monitoring Data [J]. IEEE Transactions on Instrumentation and Measurement, 2021, 70: 1-12.

[21] PENG X, LI H, YUAN F, et al. An Extreme Learning Machine for Unsupervised Online Anomaly Detection in Multivariate Time Series[J]. Neurocomputing, 2022, 501: 596-608.

[22] PANG J, PU X, LI C. A Hybrid Algorithm Incorporating Vector Quantization and One-Class Support Vector Machine for Industrial Anomaly Detection[J]. IEEE Transactions on Industrial Informatics, 2022, 18(12): 8786-8796.

[23] RUFF L, VANDERMEULEN R, GOERNITZ N, et al. Deep One-Class Classification[C]// International Conference on Machine Learning. PMLR, 2018: 4393-4402.

[24] PANG G, SHEN C, CAO L, et al. Deep Learning for Anomaly Detection: A Review[J]. ACM Computing Surveys(CSUR), 2021, 54(2): 1-38.

[25] KOPPIKAR U, SUJATHA C, PATIL P, et al. Real-World Anomaly Detection Using Deep Learning[C]// Proceedings of 3rd Intelligent Computing and Communication(ICICC 2019). Springer Singapore, 2020: 333-342.

[26] LIZNERSKI P, RUFF L, VANDERMEULEN R A, et al. Explainable Deep One-Class Classification[C]// International Conference on Learning Representations. 2021.

[27] ULGER F, YUKSEL S E, YILMAZ A. Anomaly Detection for Solder Joints Using β-VAE[J]. IEEE Transactions on Components, Packaging and Manufacturing Technology, 2021, 11(12): 2214-2221.

[28] CHEN L, LI Y, DENG X, et al. Dual Auto-Encoder Gan-Based Anomaly Detection for Industrial Control System[J]. Applied Sciences, 2022, 12(10): 4986.

[29] CHENG D, FAN Y, FANG S, et al. ResNet-AE for Radar Signal Anomaly Detection[J]. Sensors, 2022, 22(16): 6249.

[30] LI S, LIU F, JIAO L. Self-Training Multi-Sequence Learning with Transformer for Weakly Supervised Video Anomaly Detection[C]// Proceedings of The AAAI Conference on Artificial Intelligence. 2022: 1395-1403.

[31] YAN S, SHAO H, XIAO Y, et al. Hybrid Robust Convolutional Autoencoder for Unsupervised Anomaly Detection of Machine Tools Under Noises[J]. Robotics And Computer-Integrated Manufacturing, 2023, 79: 102441.

[32] TAO X, GONG X, ZHANG X, et al. Deep Learning for Unsupervised Anomaly Localization in Industrial Images: A Survey [J]. IEEE Transactions on Instrumentation and Measurement, 2022, 71: 1-21.

[33] WANG G, HAN S, DING E, et al. Student-Teacher Feature Pyramid Matching for Anomaly Detection[C]// British Machine Vision Conference. 2021.

[34] HU J, SHEN L, SUN G. Squeeze-and-Excitation Networks [C]// Proceedings of The IEEE Conference on Computer Vision and Pattern Recognition. 2018: 7132-7141.

[35] SHORTEN C, KHOSHGOFTAAR T M. A Survey on Image

Data Augmentation for Deep Learning[J]. Journal of Big Data, 2019, 6(1): 1-48.

[36] KAUR P, KHEHRA B S, MAVI E B S. Data Augmentation for Object Detection: A Review[C]// 2021 IEEE International Midwest Symposium on Circuits and Systems (MWSCAS). IEEE, 2021: 537-543.

[37] GOU J, YU B, MAYBANK S J, et al. Knowledge Distillation: A Survey[J]. International Journal of Computer Vision, 2021, 129: 1789-1819.

[38] GOLAN I, EL-YANIV R. Deep Anomaly Detection Using Geometric Transformations[J]. Advances in Neural Information Processing Systems, 2018, 31.

[39] YE F, HUANG C, CAO J, et al. Attribute Restoration Framework for Anomaly Detection[J]. IEEE Transactions on Multimedia, 2020, 24: 116-127.

[40] SCHLEGL T, SEEBÖCK P, WALDSTEIN S M, et al. Unsupervised Anomaly Detection with Generative Adversarial Networks to Guide Marker Discovery[C]// 25th International Conference Information Processing in Medical Imaging (IPMI 2017). Cham: Springer International Publishing, 2017: 146-157.

[41] NAPOLETANO P, PICCOLI F, SCHETTINI R. Anomaly Detection in Nanofibrous Materials By CNN-Based Self-Similarity [J]. Sensors, 2018, 18(1): 209.

[42] 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.

[43] YI J, YOON S. Patch SVDD: Patch-Level SVDD for Anomaly Detection and Segmentation[C]// Proceedings of the Asian Conference on Computer Vision. 2020.

[44] DEFARD T, SETKOV A, LOESCH A, et al. PaDim: A Patch Distribution Modeling Framework for Anomaly Detection and Localization[C]// Pattern Recognition. ICPR International Workshops and Challenges. Cham: Springer International Publishing, 2021: 475-489.

[45] COHEN N, HOSHEN Y. Sub-Image Anomaly Detection with Deep Pyramid Correspondences[J]. arXiv: 2005.02357, 2020.

[46] PIRNAY J, CHAI K. Inpainting transformer for anomaly detection[C]// Image Analysis and Processing (ICIAP 2022). Cham: Springer International Publishing, 2022: 394-406.

QIAO Hong, born in 1998, postgraduate. Her main research interests include anomaly detection, knowledge distillation and deep learning.

XING Hongjie, born in 1976, Ph.D, professor, Ph.D supervisor. His main research interests include kernel methods, neural networks, novelty detection, and ensemble learning.