

基于特征金字塔匹配和自监督的表面缺陷检测

梁明, 张明路, 吕晓玲*

河北工业大学机械工程学院, 天津 300401

摘要 针对传统表面缺陷检测无法适应工业复杂背景等问题, 提出一种基于特征金字塔匹配和自监督的表面缺陷检测算法。首先, 将两个基于通道注意力的残差网络提取的特征构成金字塔, 根据网络各层输出的差异找到缺陷。其次, 网络预训练的方式上采用了自我引导潜能(BYOL)自监督学习, 经过自监督学习的网络可以提取通用特征, 并提高缺陷检测方法的泛化性。最后, 在遇到模糊图像时, 采用基于不同分辨率的蒸馏训练来让学生网络充分学会提取图像的深度特征。对所提算法在 3 个数据集上进行了测试, 实验结果证明, 所提方法好于对照组, 具有更高的缺陷检测精度。

关键词 图像处理; 缺陷检测; 特征金字塔; 自监督学习; 知识蒸馏; 预训练网络

中图分类号 TP391.4

文献标志码 A

DOI: 10.3788/LOP212927

Surface-Defect Detection Based on Feature Pyramid Matching and Self-Supervision

Liang Ming, Zhang Minglu, Lü Xiaoling*

School of Mechanical Engineering, Hebei University of Technology, Tianjin 300401, China

Abstract To solve the problem of traditional surface defect detection's inadaptability to a complex industrial background, a surface-defect detection algorithm based on feature pyramid matching and self-supervision is proposed. First, the features extracted from two ResNet networks based on channel attention are structured into a pyramid, which enables defect detection using the output differences of each layer of the network. Second, bootstrap your own latent (BYOL) self-supervised learning is used in the training mode of the pretrained network; the network with self-supervised learning can extract general features and improve the generalization of the defect detection method. Finally, for fuzzy images, distillation training based on different resolutions is used to let a student network fully learn to extract the depth features of images. The proposed algorithm is tested on three datasets. Experiments show that the proposed method is better than the control group and has a higher defect detection accuracy.

Key words image processing; defect detection; feature pyramid; self-supervised learning; knowledge distillation; pre-trained network

1 引言

工业的表面缺陷检测一直是近几年来受关注较多的领域, 图片的高维数据限制了聚类等传统检测方法。在视觉领域内, 神经网络可以有效处理高维数据且速度较快^[1], 因此采用神经网络进行缺陷检测成为一种常用的解决方法。早期的表面缺陷检测采用了有监督学习, 利用标注缺陷样本在测试环节中检测出异常样本, 但是该类方法会由于工业缺陷样本不足限制训练效果, 并且碰到未标注类别的时候会产生漏检等现象。

无监督学习则成为主要的检测方式。传统无监督缺陷检测是基于图像预处理方法^[2-3]或者深度网络提取特征^[4-5]来放入支持向量机进行缺陷检测, 但是该类方法不能满足工业复杂数据的检测。随后的无监督缺陷检测大多基于自编码器^[6-7]或生成对抗网络(GAN)^[8-9]等, 这些方法都是基于重构误差来检测出缺陷样本, 但是如果训练的编码器效果过好, 会产生一种误差失灵的现象。为了放大正常样本与缺陷样本的差异, 这一类基于知识蒸馏的检测方法活跃在缺陷检测领域内。文献[10]提出在无异常的样本上对教师学生网络进行

收稿日期: 2021-11-11; 修回日期: 2021-12-06; 录用日期: 2021-12-21; 网络首发日期: 2021-12-30

基金项目: 国家重点研发计划(2017YFB1303700)

通信作者: *lx1000418@163.com

蒸馏学习,测试中将各网络的输出聚类,把没有聚成一个类簇的样本定义为缺陷样本。文献[11]提出将蒸馏学习输出的过程改为对中间层特征的学习,测试环节中利用可解释性算法查找出令其损失梯度变化较大的地方,将其定义为缺陷区域。文献[12-13]为了提升方法的缺陷检测速度,均利用了各网络提取的特征构成金字塔来进行快速检测。

基于知识蒸馏的缺陷检测方法较以往方法精度更高,但是该类方法都是在基于已有预训练网络的基础上进行检测的,网络的结构会受到限制,并且当目标领域与预训练网络的源领域差异较大时,编码器提取特征的能力会大打折扣。其次该方法的蒸馏训练是在

相同分辨率的样本上进行的,在遇到图像模糊的情况下,会因分辨率过小导致检测效果受到影响。本文针对上述问题,在原始的预训练网络中加入通道注意力机制来提高网络提取特征的能力,框架中教师网络的预训练模式上添加了自监督预训练方式,并在检测模糊图像的领域时采用基于不同分辨率的蒸馏训练来提高方法异常检测的性能。

2 本文方法

2.1 方法框架

本文方法的总体框架如图 1 所示,分为 3 个部分:网络预训练部分、蒸馏训练部分和缺陷检测部分。

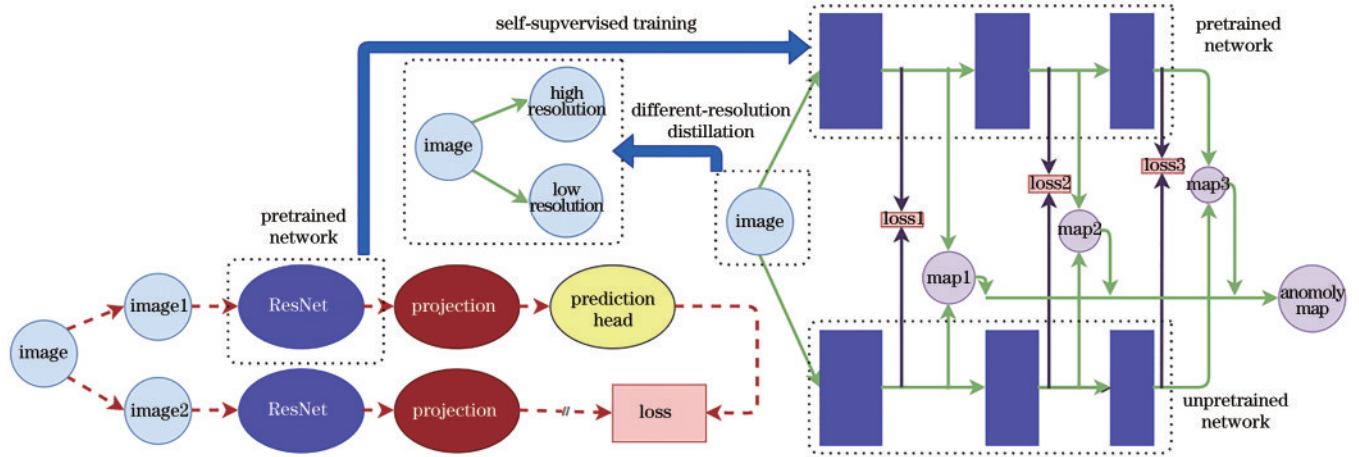


图 1 基于特征金字塔匹配和自监督的缺陷检测框架图

Fig. 1 Frame diagram of defect detection based on feature pyramid matching and self-supervision

网络预训练部分采用自监督学习进行预训练,如图 1 左侧将数据通过孪生残差网络进行对比学习,训练完成后保留编码器的权重并导入图 1 右侧蒸馏框架的教师网络中,以供后续蒸馏部分的训练,自监督学习可以提高网络的泛化性,避免目标领域与源数据领域差异较大带来的影响。

蒸馏训练部分采用的是两个基于通道注意力的 ResNet50 孪生网络,通过网络各层输出特征图的相似计算,将未训练的学生网络在正常样本上对预训练过的教师网络进行蒸馏学习。蒸馏训练的方式采用基于不同分辨率的蒸馏学习,将样本原图像输入教师网络,不同分辨率的图像输入学生网络,通过该方式让方法在目标领域图像较为模糊的情况下,提高对含有整体信息深层特征的提取能力。

缺陷检测部分以蒸馏训练后的教师学生网络为基础,测试时,当图像输入网络框架时,通过网络各层的差异为图像中的每像素点设置得分,用来表达与正常样本像素点偏离的程度。各层得出的分数图呈金字塔形状大小不一,需要线性插值得出具有原图大小的异常分数图,最后异常分数图通过高斯模糊来显现异常区域的边缘,并设置阈值来圈出缺陷位置。

2.2 通道自注意力机制模块

本文对常规的网络加入通道自注意力(SE)模块,以此来提升网络的编码能力。SE 模块如图 2 所示,该模块通过池化层利用式(1)将输入特征图的空间维度进行压缩,将二维的特征通道变成一个拥有全局感受野的实数。

$$z_c = F_s(u_c) = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W u_c(i, j), \quad (1)$$

式中: u_c 为原始输出的特征图; F_s 为全局平均池化操作; W 和 H 分别为特征图的宽与高; z_c 为池化后的特征图。

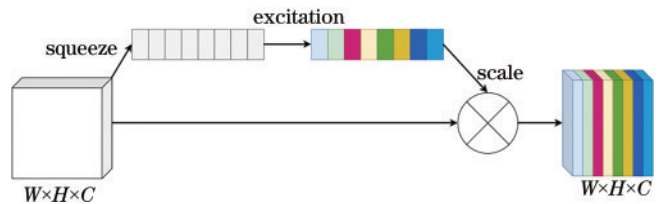


图 2 基于通道注意力的模块示意图

Fig. 2 Schematic diagram of module based on channel attention

随后通过式(2)获得全面的通道级别的权重,最后为输入相应的通道附加实数权重,利用式(3)将权重与原始的特征融合,完成对原始特征的重新标定。

$$s_c = \sigma [W_2 \sigma (W_1 z_c)], \quad (2)$$

$$X = s_c \times u_c, \quad (3)$$

式中： σ 为激活函数； W_1 和 W_2 为全连接层； X 为SE模块最后输出的特征图。

SE机制使每个通道都拥有一个在下一阶段中的重要程度信息，可以更好地拟合相关通道复杂关系，并且可以减少网络参数数量和计算量。

如图3所示，左侧为残差网络的常规残差块，右侧为加入SE模块后的残差块。本文在网络预训练时将SE模块插入ResNet50的每一个残差块中，利用该模块使网络学到各通道的关系并赋予相应的权重，提升网络提取特征的能力。

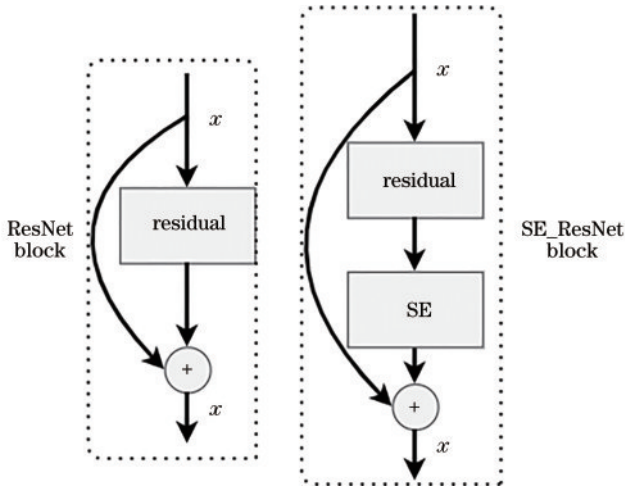


图3 通道注意力模块插入方式

Fig. 3 Insertion mode of channel attention module

2.3 基于特征金字塔蒸馏训练

为了适应工业快速检测的要求，本文借鉴文献[12]利用特征金字塔在线匹配出缺陷图像。网络训练时对各层提取的特征图做相似计算，利用L2损失计算两特征图对应像素点位置上特征向量的相似度。具体如下式所示：

$$\left\{ \begin{aligned} f_t^k(I_a)_{ij} &= \frac{f_t^k(I_a)_{ij}}{\|f_t^k(I_a)_{ij}\|_{12}} \\ f_s^k(I_a)_{ij} &= \frac{f_s^k(I_a)_{ij}}{\|f_s^k(I_a)_{ij}\|_{12}} \\ l(I_a)_{ij}^k &= \frac{1}{2} \|f_t^k(I_a)_{ij} - f_s^k(I_a)_{ij}\|_{12}^2, \\ l_a &= \sum_{k=1}^3 \left(\frac{1}{w_a \times h_a} \sum_{i=1}^{w_a} \sum_{j=1}^{h_a} l(I_a)_{ij}^k \right) \\ l_{\text{total}} &= \sum_{a=1}^n l_a \end{aligned} \right. \quad (4)$$

式中： w 和 h 分别为输入图像的宽和高； k 为网络的层次编号； a 为图像的编号； n 为图像的总数； s 代表学生

网络； t 代表教师网络； i, j 分别为图像像素点的位置坐标； total 代表总损失。

测试时将测试集的图片同时输入教师学生网络，两者网络各层提取的特征图构成金字塔。通过两者差异为图中各像素点设置相应的得分 $S \in (0, 1)$ ，分数用来表示该图像像素点与正常样本像素点偏离的程度。于是教师学生网络各层会给出不同尺度的异常分数图，最后利用双线性插值技术式(5)得出具有原图尺寸的异常分数图。

$$S(I_a) = \prod_k^3 u[S^k(I_a)], \quad (5)$$

式中： k 为网络的层次编号； I 为输入的图像； $u()$ 为双线性插值函数。

本文考虑到背景和噪声对缺陷区域提取的干扰，对最终的异常分数图进行高斯模糊，高斯模糊不仅可以去噪，还能加大图像深浅处的区别，使缺陷检测的精度提高，最后在异常分数图上通过设置阈值来标出缺陷区域。

2.4 基于不同分辨率的蒸馏训练

虽然大部分数据集的图像分辨率较高，但是实际的工业环境下，有些图像往往都是又小又模糊。模糊图像的识别分类一般都是依靠含有整体信息的深层特征。如果只将训练集中的同一样本输入教师学生网络，由于蒸馏学习存留的差异性，学生网络往往做不到完全像教师网络那样抽象化提取特征，会让方法局限于该训练集的分辨率中。考虑蒸馏学习应该摆脱分辨率的影响，如图4所示，本节提出在教师网络中输入原图像，在学生网络中输入不同分辨率的图像，该图像通过原图像下采样并插值回原图的大小来获得。

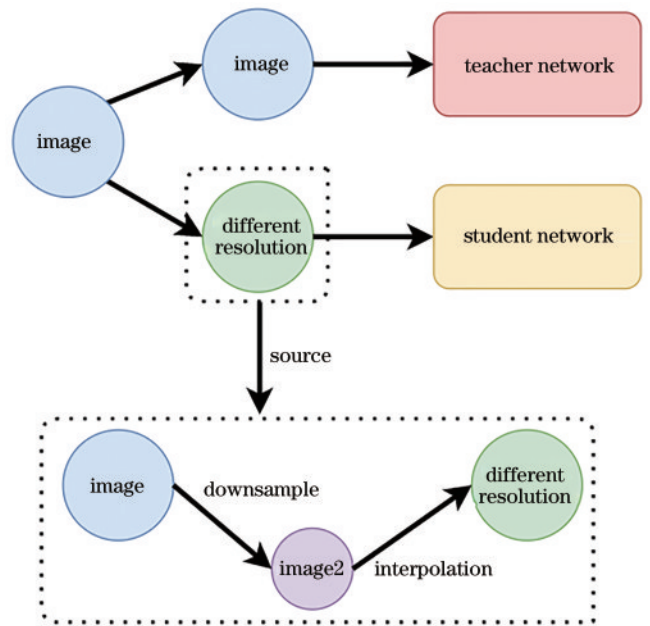


图4 基于不同分辨率的蒸馏学习

Fig. 4 Distillation learning based on different resolutions

在基于不同分辨率图像的蒸馏训练时,网络生成的两个分辨率的特征图会被不断拉近,使学生网络更加充分地学习到教师网络在模糊图像上的抽象提取方式,提高本文方法在原分辨率较小图片上的异常检测效果。

2.5 自监督预训练方式

网络的训练方式有很多种,以往缺陷检测方法的网络都是采用已有预训练网络或者自己标注样本的训练网络。然而已有的预训练网络会有很大的局限性,该网络大多是最原始的网络,如果想对网络的结构进行更改,已有的网络权重将不再适用。并且当源数据集与目标数据集有很大的差异时,预训练网络反而会起到反作用,提取到对下游任务不利的特征,影响检测效果。自监督式的训练网络虽然可以随意更改网络结构并且缓解泛化性差等现象,但是有监督学习需要大量标注的数据,标注数据环节需要耗费大量时间和人力,并且缺陷样本匮乏的工业背景也成了难题,于是不少学者对扩充样本开展了研究。对于样本不足问题,文献[14-15]采用神经网络进行重构再生成,但深度网络难以训练,参数过多。文献[16]虽提出利用GAN来扩充数据集,但是GAN的训练不稳定且计算量复杂,因此按照缺陷样本进行网络预训练是一个困难的方法。

自监督学习为缺陷检测网络训练中出现的问题带来了解决方法。自监督学习可以通过大量无标注数据训练出一个强大的特征编码器来提取到通用特征,并将学好的权重在下游任务上微调即可。近期已有多种实验结果表明自监督预训练好于有监督预训练网络。自我引导潜能(BYOL)^[17]算法的出现是自MoCo^[18]、SimCLR^[19]算法后的又一大里程碑,BYOL自监督算法在提出不用负样本对的情况下,达到了当时SOTA的效果。该算法框架如图5所示,算法模型由两个相同的编码器框架组成,编码器含有网络结构和映射头。图5左侧框架比右侧框架多出一个起到关键作用的预测头。预测头可以为方法带来不对称的作用,从而充当了负样本对,防止模型在训练时崩溃。整体框架的训练就是计算同一图像的不同视角下的相似度,拉近产自同一图像的样本的距离,推远产自不同图片的样本的距离。左侧框架采用梯度传播更新编码器中的参数,右侧编码器的参数通过左侧框架的编码器进行动量编码更新。摆脱负样本对的想法使得算法对于批次的大小不再敏感,性价比也好于同等参数情况下的其他算法。

本文在进行网络自监督训练时,利用无标签数据进行训练。训练完成后,保留编码器的权重。随后将编码器中网络的权重导入教师网络,以供后续蒸馏学习和缺陷检测等操作。

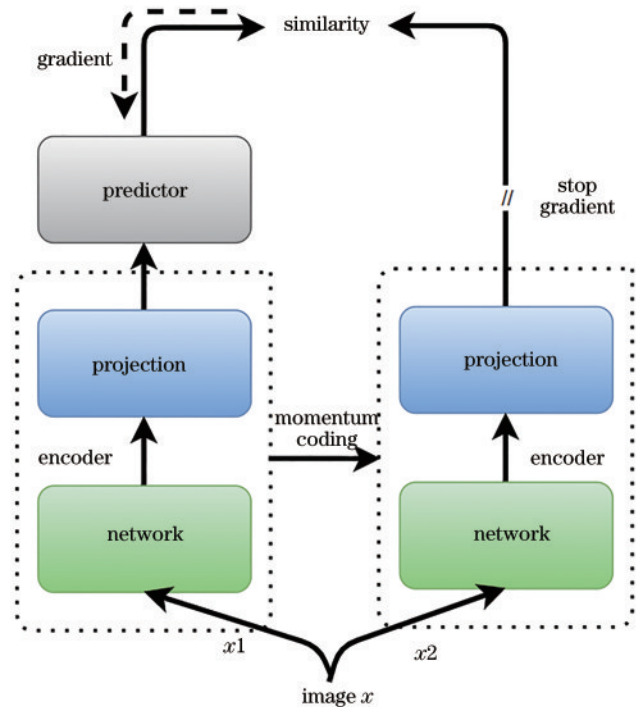


图5 BYOL算法框架图

Fig. 5 Frame diagram of BYOL algorithm

3 实验结果与分析

3.1 数据集与评估指标

本文测试采用的是MVTec AD^[20]数据集、CIFAR-10^[21]数据集和STL-10^[22]数据集。MVTec AD数据集主要用来检测算法的异常检测和缺陷定位的性能。后两个数据集由于拥有大量无标注的数据,用来检测自监督预训练方式对检测方法带来的影响,同时较为模糊的CIFAR-10数据集用来检测基于不同分辨率的蒸馏训练对检测方法带来的影响。

MVTec AD数据集是近几年发布的具有挑战性的工业产品无监督缺陷检测数据集。数据集内一共包含15个类别。训练集只包括正常样本,共有3629张图片。测试集包含正常样本和缺陷样本,缺陷样本模拟了工业中产品出现的各种缺陷,共有1725张图片。CIFAR-10数据集是由60000张32 pixel×32 pixel的彩色图片组成的,数据集涵盖10个类别,每种类别有6000张图片。训练集由50000张图片用来做训练,测试集由10000张图片用来做测试。所有图片清晰度较低,分辨率较小。STL-10数据集是一个专门用于研究无监督特征学习的图像识别数据集。训练集含有100000张无标签图像,以供无监督学习。测试集含有10个类别,共5000张有标签图像。该数据集主要的挑战是利用未标记的数据(与标记数据相似但分布不同)来构建有用的先验知识。

本文评估指标与相关的无监督缺陷检测方法所用

指标类似,采用的是可以通过设置阈值来检测的接收器工作特性下的面积(AUROC)指标,本文主要在该指标图像级异常检测和像素级缺陷定位的两个工作进行评估,其次为了检验算法的可靠性和检测速度,分别给出召回率、误检率、漏检率和图片检测时间等指标上的结果。

3.2 实验结果与分析

为了验证本文算法的有效性,分别以经典的SSIM-AE^[23]、Anogan^[9]、KD-AD^[11]、SPADE^[13]和本文方法在MVTec AD数据集上进行实验,为了方便实验,本文方法利用已有预训练的SE_ResNet50网络作为框架基础进行检测,传统SSIM-AE和Anogan方法的结果可查看文献原有的结果,其余近期方法通过复现给出结果,为了保证复现效果最优,利用官方源码按照文献提供的训练参数进行复现,得出结果与原文最优结果相差不大,表1~2分别列出了复现方法与本文方法检测结果的对比。

根据表1实验结果可知,在异常检测的指标上,本文方法比起SPADE和KD-AD方法分别有10.59个百分点和8.10个百分点的提升,在缺陷定位的指标上,分别有1.74个百分点和6.89个百分点的

提升,并且比原文中SSIM-AE和Anogan方法的结果有更高的提升。由表2结果可知,本文方法在误检率上比起SPADE和KD-AD方法分别低6.88个百分点和3.78个百分点,在漏检率上分别低5.64个百分点和5.75个百分点,召回率达到97.68%。从表1~2中还可以看到,在本文方法中加入了通道注意力机制的网络比起未加入的网络方法检测效果提升明显,异常检测指标提高了1.11个百分点,缺陷定位指标有0.82个百分点的提升,误检率和漏检率分别低0.69个百分点和0.44个百分点。从图片的检测时间可知,对于每一个256 pixel×256 pixel的图片,通道注意力机制模块的加入并没有对本文算法的缺陷检测时间带来很大增加,只需要35 ms,远低于对比方法的检测时间,这是因为SPADE方法需要笨重的Wide-ResNet50×2网络来提取特征,双步骤的检测也会增加计算量,严重拖慢检测速度,KD-AD方法则在网络进行正向传播后还要再进行一次反向传播梯度来检测缺陷,本文方法采用轻便的残差网络,且只需要通过网络一次正向传播,根据各层差异检测出缺陷,由实验结果可知,本文方法能够快速检测出物体的表面缺陷。

表1 MVTec AD数据集上的检测结果
Table 1 Test results on MVTec AD dataset

Parameter	Image AUROC / %				Pixel AUROC / %			
	SPADE	KD-AD	Proposed	Proposed (SE)	SPADE	KD-AD	Proposed	Proposed (SE)
Leather	94.47	95.26	100.00	100.00	96.94	98.15	99.24	99.53
Grid	75.36	78.52	94.62	97.25	93.72	91.78	99.00	98.16
Bottle	98.27	99.45	100.00	100.00	98.13	96.19	98.44	98.76
Transistor	82.53	85.47	95.96	94.42	93.98	76.04	89.26	89.18
Tile	88.68	91.63	98.99	100.00	87.47	82.31	96.53	95.95
Pill	80.24	83.30	94.93	97.77	96.44	90.08	96.02	97.71
Wood	91.39	94.51	99.04	99.72	88.26	84.65	95.90	96.37
Cable	85.42	89.38	98.20	97.23	97.05	82.59	97.38	97.70
Capsule	82.15	80.72	84.00	86.57	98.79	95.42	97.25	98.59
Carpet	77.04	80.05	97.03	95.04	96.82	95.74	98.98	99.22
Toothbrush	88.89	92.49	88.06	90.70	97.98	96.25	98.35	98.50
Zipper	90.12	93.58	95.01	94.67	96.48	94.24	97.27	98.03
Hazelnut	95.85	98.56	98.14	100.00	98.52	94.79	98.60	98.80
Metal_nut	71.94	73.74	89.15	98.15	98.21	86.25	92.41	98.30
Screw	80.86	83.95	92.29	90.54	98.81	95.84	96.79	98.90
Average	85.55	88.04	95.03	96.14	95.84	90.69	96.76	97.58

表2 MVTec AD数据集上其他指标检测结果
Table 2 Test results of other indicators on MVTec AD dataset

Parameter	Recall / %	False rate / %	Miss rate / %	Time / ms
SPADE	92.04	10.72	7.96	1530
KD-AD	91.93	7.62	8.07	84
Proposed	97.24	4.53	2.76	33
Proposed (SE)	97.68	3.84	2.32	35

在STL-10和CIFAR-10数据集上,基于多种网络进行对比检验不同分辨率蒸馏训练方式和自监督预训练方式对本文方法的影响,随后在SE_ResNet50网络基础上逐个添加训练方法模块,进行消融实验。表3~5分别列出了在两个数据集上缺陷检测结果的对比,表格中用resolution代表采用基于不同分辨率蒸馏训练方式,BYOL代表采用基于BYOL的自监督预训练方式。由表3~4的实验结果可知,虽然在MVTec AD

表 3 STL-10数据集上的检测结果
Table 3 Test results on STL-10 dataset

Parameter	Image AUROC / %					
	ResNet18	ResNet50	SE_ResNet50	ResNet18+ BYOL	ResNet50+ BYOL	SE_ResNet50+ BYOL
Truck	82.53	82.78	80.62	85.84	87.21	85.96
Horse	79.08	79.30	76.32	83.14	82.95	80.21
Deer	86.05	87.75	88.12	85.90	89.30	90.47
Car	86.93	89.35	89.93	90.38	92.48	92.78
Cat	73.16	73.89	76.89	77.75	77.52	79.53
Ship	92.38	93.47	92.97	93.74	94.08	93.62
Dog	72.45	73.81	72.78	69.84	70.94	68.48
Airplane	91.11	91.40	89.81	90.13	89.99	87.70
Monkey	79.73	82.69	83.86	80.20	83.57	84.62
Bird	77.09	79.23	78.47	77.85	80.15	78.85
Average	82.05	83.37	82.98	83.48	84.82	84.22

表 4 CIFAR-10数据集上的检测结果
Table 4 Test results on CIFAR-10 dataset

Parameter	Image AUROC / %										
	0	1	2	3	4	5	6	7	8	9	Average
ResNet18	85.28	79.62	70.95	71.53	82.75	80.97	86.35	86.74	90.27	80.43	81.49
ResNet50	86.08	79.89	70.75	70.06	87.17	81.32	83.99	90.21	88.86	78.88	81.72
SE_ResNet50	83.45	77.25	69.86	73.95	89.45	77.34	87.72	89.82	85.38	76.96	81.12
ResNet18+resolution	86.69	81.57	70.57	73.18	87.56	80.05	88.71	85.86	89.02	81.69	82.49
ResNet50+resolution	86.56	80.92	70.48	73.35	88.29	79.97	86.08	87.99	88.24	80.14	82.20
SE_ResNet50+resolution	84.36	78.59	69.45	75.21	90.27	76.21	90.54	87.58	84.97	78.52	81.57
ResNet18+BYOL	85.38	81.65	68.19	73.26	87.28	83.37	87.36	91.32	90.73	83.14	83.17
ResNet50+BYOL	86.82	81.74	69.62	73.83	90.16	84.02	85.18	92.46	88.95	79.53	83.23
SE_ResNet50+BYOL	84.08	79.74	68.92	75.44	90.89	79.85	90.21	91.53	86.23	79.21	82.61

数据集上,加入通道注意力机制模块的网络有明显的效果提升,而在 CIFAR-10 和 STL-10 这种异常样本与正常样本本质不同但多样化的数据集上,通道注意力机制并没有带来检测精度的提升,而经过自监督预训练的方法却比利用已有预训练网络的方法的异常检测精度提升明显,普遍提高约 1.5 个百分点。分析可知,STL-10 数据集与 CIFAR-10 数据集是 ImageNet 数据集的子集,分辨率分别缩小至 96 pixel×96 pixel 和 32 pixel×32 pixel,这种细微的改变就已经局限了预训练

网络的编码能力,基于自监督学习的网络由于是在该分辨率训练集下进行的训练,所以可以避免这种影响。在表 4 的 CIFAR-10 这种分辨率较小的图像数据集检测结果可知,本文在各种残差网络的基础上进行了基于不同分辨率的蒸馏训练,可以看到异常检测的性能被普遍带来了约 0.5 到 1.0 个百分点的提升。这可以证明该蒸馏训练方法能够使网络更加注重学到含有整体信息的深度特征,并对工业上这种小且模糊图像的异常检测起到积极作用。从表 5 的消融实验中可以看

表 5 STL-10 和 CIFAR-10 数据集上其他指标检测结果
Table 5 Test results of other indicators on STL-10 and CIFAR-10 datasets

Parameter	STL-10				CIFAR-10			
	Recall / %	False rate / %	Miss rate / %	Time / ms	Recall / %	False rate / %	Miss rate / %	Time / ms
SE_ResNet50	91.21	14.35	8.79	34	90.02	15.58	9.98	34
SE_ResNet50+resolution	91.38	14.22	8.62	34	90.40	15.04	9.60	35
SE_ResNet50+BYOL	92.16	13.12	7.84	35	91.08	14.52	8.92	34
SE_ResNet50+resolution+BYOL	92.28	12.98	7.72	35	91.44	14.18	8.56	35

出,基于BYOL自监督预训练比基于不同分辨率蒸馏训练带来的提升更加明显。在平均误检率和漏检率上,同时利用两训练方法比未经过任何训练的方法分别低 1.39 个百分点和 1.25 个百分点,平均召回率达到 91.86%。从图片检测时间可知,这些训练方法并没有影响到下游缺陷检测任务的检测速度,使本文算法仍然具有快速检测缺陷的优势。

3.3 测试结果展示

本节列出在 MVTEC AD 数据集上一些样例的缺陷定位检测结果。随机选取了 5 个样例,分别是皮革、瓷砖、胶囊、木头和金属螺母。图 6(a)均为原始缺陷图片,图 6(b)均为真实的缺陷标签图,图 6(c)均为本文方法检测出的异常分数图,图 6(d)均为本文方法检测出的缺陷定位结果图。

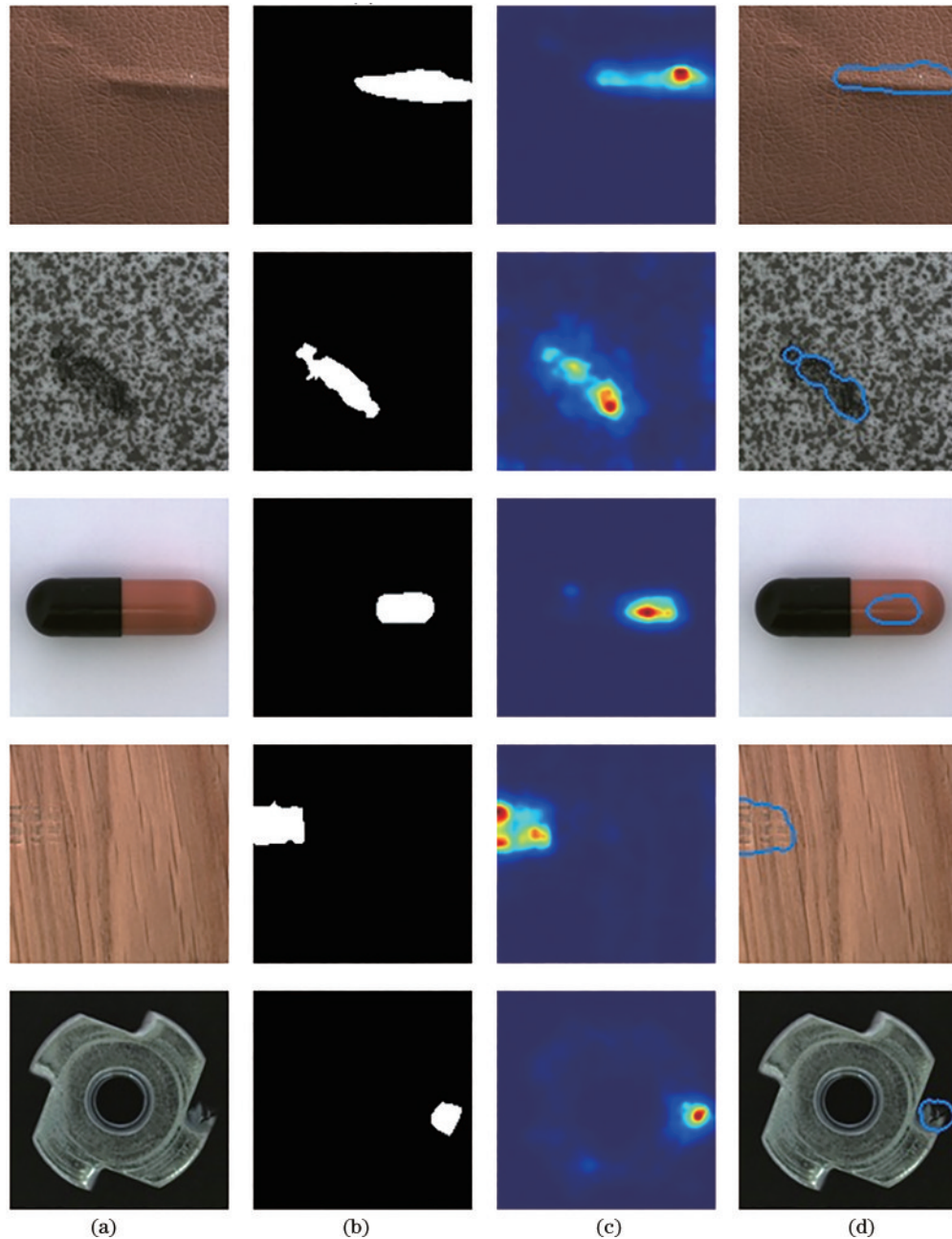


图 6 MVTEC AD 数据集上一些样例的检测结果。(a) 原图;(b) 真实标签;(c) 异常分数图;(d) 缺陷图
Fig. 6 Test results of some samples on MVTEC AD dataset. (a) Original images; (b) ground truth; (c) abnormal score maps; (d) defective images

由图 6 可以看出,本文方法所得出的异常分数图可以形象地表达出某些区域是否为缺陷的可能性,且通过在异常分数图上设置阈值可以精确地在原图中呈现出缺陷区域。

4 结 论

提出了一种基于特征金字塔匹配和自监督的缺陷检测方法,能够在工业背景下快速检测到物体表面的

缺陷。首先,将通道注意力机制引入教师学生各常规网络中,利用网络各层提取的特征构成金字塔,以便于快速检测出异常区域。此外,考虑预训练网络结构不可改变的局限性和实际的工业背景,提出引入自监督预训练方式。然后,面对分辨率不足的图像时,提出了基于不同分辨率的蒸馏学习方式。最后,在 3 个数据集上的实验结果表明,提出的检测方法优于对照组;但该方法所用的网络还有一定的改进空间,如何设置有利于下游缺陷检测任务的网络成为今后的研究方向。

参 考 文 献

- [1] Xu L, Rahmani M, Ma Y X, et al. Enhanced light-matter interactions in dielectric nanostructures via machine-learning approach[J]. *Advanced Photonics*, 2020, 2(2): 026003.
- [2] 李丹, 金媛媛, 童艳, 等. 基于支持向量机的输液袋智能检测与缺陷分类[J]. *激光与光电子学进展*, 2019, 56(13): 131502.
Li D, Jin Y Y, Tong Y, et al. Intelligent detection and defect classification of infusion bags based on support vector machine[J]. *Laser & Optoelectronics Progress*, 2019, 56(13): 131502.
- [3] 王泽润, 方益明, 冯海林, 等. 木材节子缺陷检测与定位方法[J]. *激光与光电子学进展*, 2018, 55(5): 051501.
Wang Z R, Fang Y M, Feng H L, et al. Method for wooden knot detection and localization[J]. *Laser & Optoelectronics Progress*, 2018, 55(5): 051501.
- [4] Defard T, Setkov A, Loesch A, et al. PaDiM: a patch distribution modeling framework for anomaly detection and localization[M]//del Bimbo A, Cucchiara R, Sclaroff S, et al. *Pattern recognition. ICPR international workshops and challenges. Lecture notes in computer science*. Cham: Springer, 2021, 12664: 475-489.
- [5] Burlina P, Joshi N, Wang I J. Where's wally now? deep generative and discriminative embeddings for novelty 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: 11499-11508.
- [6] Mei S, Yang H, Yin Z P. An unsupervised-learning-based approach for automated defect inspection on textured surfaces[J]. *IEEE Transactions on Instrumentation and Measurement*, 2018, 67(6): 1266-1277.
- [7] 薛英杰, 陈颀, 周松斌, 等. 基于自监督特征提取的机械异常声音检测研究[J]. *激光与光电子学进展*, 2022, 59(12): 1215013.
Xue Y J, Chen Q, Zhou S B, et al. Research on mechanical abnormal sound detection based on self supervised feature extraction[J]. *Laser & Optoelectronics Progress*, 2022, 59(12): 1215013.
- [8] Schlegl T, Seeböck P, Waldstein S M, et al. F-AnoGAN: fast unsupervised anomaly detection with generative adversarial networks[J]. *Medical Image Analysis*, 2019, 54: 30-44.
- [9] 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.
- [10] Bergmann P, Fauser M, Sattlegger D, et al. Uninformed students: student-teacher anomaly detection with discriminative latent embeddings[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 13-19, 2020, Seattle, WA, USA. New York: IEEE Press, 2020: 4182-4191.
- [11] Salehi M, Sadjadi N, Baselizadeh S, et al. Multiresolution knowledge distillation for anomaly detection[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 20-25, 2021, Nashville, TN, USA. New York: IEEE Press, 2021: 14897-14907.
- [12] Wang G D, Han S M, Ding E R, et al. Student-teacher feature pyramid matching for anomaly detection[EB/OL]. (2021-03-07) [2021-06-07]. <https://arxiv.org/abs/2103.04257>.
- [13] Cohen N, Hoshen Y. Sub-image anomaly detection with deep pyramid correspondences[EB/OL]. (2020-05-09) [2021-11-06]. <https://arxiv.org/abs/2005.02357>.
- [14] Dong D S, Shi K B. Solving the missing cone problem by deep learning[J]. *Advanced Photonics*, 2020, 2(2): 020501.
- [15] Lim J, Ayoub A B, Psaltis D. Three-dimensional tomography of red blood cells using deep learning[J]. *Advanced Photonics*, 2020, 2(2): 026001.
- [16] 常江, 管声启, 师红宇, 等. 基于改进生成对抗网络和 MobileNetV3 的带钢缺陷分类[J]. *激光与光电子学进展*, 2021, 58(4): 0410016.
Chang J, Guan S Q, Shi H Y, et al. Strip defect classification based on improved generative adversarial networks and MobileNetV3[J]. *Laser & Optoelectronics Progress*, 2021, 58(4): 0410016.
- [17] Grill J B, Strub F, Alché F, et al. Bootstrap your own latent: a new approach to self-supervised learning[EB/OL]. (2020-06-13) [2021-04-05]. <https://arxiv.org/abs/2006.07733>.
- [18] He K M, Fan H Q, Wu Y X, et al. Momentum contrast for unsupervised visual representation learning[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 13-19, 2020, Seattle, WA, USA. New York: IEEE Press, 2020: 9726-9735.
- [19] Chen T, Kornblith S, Norouzi M, et al. A simple framework for contrastive learning of visual representations [C]//ICML'20: Proceedings of the 37th International Conference on Machine Learning, July 13-18, 2020, Virtual. New York: ACM Press, 2020: 1597-1607.
- [20] Bergmann P, Batzner K, Fauser M, et al. The MVTec anomaly detection dataset: a comprehensive real-world dataset for unsupervised anomaly detection[J]. *International Journal of Computer Vision*, 2021, 129(4): 1038-1059.
- [21] Krizhevsky A, Nair V, Hinton G. Learning multiple

- layers of features from tiny images[EB/OL]. (2009-04-08) [2021-11-11]. <https://www.semanticscholar.org/paper/Learning-Multiple-Layers-of-Features-from-Tiny-Krizhevsky/5d90f06bb70a0a3dced62413346235c02b1aa086>.
- [22] Coates A, Lee H, Andrew Y. An analysis of single layer networks in unsupervised feature learning[C]//Proceedings of the International Conference on Artificial Intelligence and Statistics, April 11-13, 2011, Fort Lauderdale, USA. Cambridge: JMLR, 2011: 215-223.
- [23] Bergmann P, Lowe S, Fauser M, et al. Improving unsupervised defect segmentation by applying structural similarity to autoencoders[C]//International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP), February 25-27, 2019, Prague, Czech Republic. Valletta: SciTePress, 2019: 372-380.