计算机科学
COMPUTER SCIENCE

面向工业图像异常检测的非对称师生网络模型

孔森林, 张辉, 黄镇南, 刘优武, 陶岩

引用本文

孔森林, 张辉, 黄镇南, 刘优武, 陶岩. 面向工业图像异常检测的非对称师生网络模型[J]. 计算机科学, 2024, 51(11A): 240200069-7.

KONG Senlin, ZHANG Hui, HUANG Zhennan, LIU Youwu, TAO Yan. Asymmetric Teacher-Student Network Model for Industrial Image Anomaly Detection [J]. Computer Science, 2024, 51(11A): 240200069-7.

计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.240200069

面向工业图像异常检测的非对称师生网络模型

孔森林[1]　张　辉[2]　黄镇南[3]　刘优武[1]　陶　岩[1]
1 长沙理工大学电气与信息工程学院　长沙 410000
2 湖南大学机器人学院　长沙 410000
3 中国人民武装警察部队警官学院　成都 610213
(986735244@qq.com)

摘　要　工业图像异常检测是大规模工业制造中的关键组成部分。针对工业图像异常检测存在的异常样本标注难度大、异常区域先验信息获取困难等问题,提出了一种基于非对称师生网络的无监督图像异常检测模型。首先,针对高相似结构师生网络导致的过模仿映射问题,设计了非对称师生网络,通过向学生网络残差块中引入上下文 Transformer 模块,为师生网络添加结构差异性,阻止学生网络过模仿教师网络的映射。其次,为了增强师生网络之间的泛化性差异,在教师网络中引入移动平均归一化层,以提高检测性能。最后,引入多尺度异常图融合机制,通过融合不同尺度的异常分数图,以更好地检测不同大小的异常。在 MVTec AD 公共数据集上进行了相关实验,实验结果中图像级别 AUROC 达到 95.7%,像素级别 AUROC 达到97.4%,验证了该方法的可行性和有效性。

关键词:异常检测;知识蒸馏;Transformer;无监督学习;多尺度特征
中图分类号　TP391

Asymmetric Teacher-Student Network Model for Industrial Image Anomaly Detection

KONG Senlin[1],ZHANG Hui[2],HUANG Zhennan[3],LIU Youwu[1] and TAO Yan[1]
1 School of Electrical & Information Engineering,Changsha University of Science and Technology,Changsha 410000,China
2 School of Robotics,Hunan University,Changsha 410000,China
3 Officers College of PAP,Chengdu 610213,China

Abstract　Industrial image anomaly detection is a critical component in large-scale industrial manufacturing. Addressing challenges such as difficulty in annotating anomalous samples and obtaining prior information about anomalous regions in industrial image anomaly detection,a model based on asymmetric teacher-student networks for unsupervised image anomaly detection is proposed. Firstly,to tackle the problem of over-imitation mapping caused by high similarity in structure between teacher and student networks,an asymmetric teacher-student network is designed. Contextual Transformer modules are introduced into the residual blocks of the student network to add structural diversity to the teacher-student networks,preventing the student network from over-imitating the mapping of the teacher network. Secondly,to enhance the generalization difference between teacher and student networks,a moving average normalization layer is introduced into the teacher network to improve detection performance. Finally,a multi-scale abnormality map fusion mechanism is introduced to better detect anomalies of different sizes by fusing abnormality score maps of different scales. Experiments conducted on the MVTec AD public dataset show that the proposed method achieves an image-level AUROC of 95.7% and a pixel-level AUROC of 97.4%,verifying the feasibility and effectiveness of the approach.

Keywords　Anomaly detection,Knowledge distillation,Transformer,Unsupervised learning,Multi-scale features

---

基金项目:科技创新 2030—“新一代人工智能”重大项目(2021ZD0114503);国家自然科学基金重大研究计划(92148204);国家自然科学基金(62027810);湖南省科技创新领军人才(2022RC3063);湖南省杰出青年科学基金项目(2021JJ10025);湖南省重点研发计划(2021GK4011,2022GK2011);长沙科技重大项目(KH2003026);中国高校产学研创新基金(2020HYA06006);湖南省研究生科研创新项目(CX20220923);长沙理工大学研究生科研创新项目(CXCLY2022088)
This work was supported by the Science and Technology Innovation 2030 — “New Generation Artificial Intelligence” Major Project (2021ZD0114503),National Natural Science Foundation of China Major Research Program(92148204),National Natural Science Foundation of China(62027810),Leading Scientific and Technological Innovation Talents of Hunan Province(2022RC3063),Hunan Outstanding Young People Science Foundation Project(2021JJ10025),Hunan Key Research and Development Project(2021GK4011,2022GK2011),Changsha Key Science and Technology Project(KH2003026),China University Industry University Research Innovation Fund (2020HYA06006),Hunan Graduate Research Innovation Project (CX20220923) and Changsha University of Science and Technology Graduate Research Innovation Project (CXCLY20222088).
通信作者:张辉(zhanghuihby@126.com)

1 引言

在工业生产的过程中，确保产品质量的稳定性和一致性是维护生产安全和可持续发展的关键要素之一[1]。然而，由于生产环境具有复杂性和多变性，工业产品在制造过程中很容易受到各种因素的影响，例如材料缺陷、工艺偏差、设备故障等，导致产品可能含有各种各样的异常。这些异常不仅对产品质量造成直接影响，而且可能增加不必要的生产成本并导致资源浪费。所以，对产品进行异常检测是工业制造过程不可或缺的环节之一。

为了有效地识别工业产品表面的异常和缺陷，传统图像检测方法通常采用手工设计的特征，如采用特征点[2-3]、局部边缘特征[4-5]和局部区域特征[6-8]进行匹配。然而，这些方法过度依赖人工，往往需要大量专业知识和经验来选择和提取特征。此外，它们的泛化能力受到特征设计的限制，使其在面对光照、尺度、变形等变化时表现出一定的敏感性。在复杂多变的生产环境中，传统图像检测方法还面临着图像噪声的挑战。

随着深度学习技术的迅猛发展，尤其是卷积神经网络模型的成功应用，深度学习在工业产品异常检测领域展现出了巨大的潜力[9]。有监督深度学习方法，如双阶段的 RCNN 系列算法[10-11]、单阶段的 YOLO 算法[12-13]和 SSD 算法[14-15]，已经在异常检测领域取得了显著的发展和应用。

然而，基于有监督的深度学习方法通常需要对大量异常数据集进行标注，标记数据的收集和准备成本往往较高[16]。这种情况下，模型的性能高度依赖于标签数据的质量，不准确或不完整的数据标注可能导致不良结果。而且，在实际生产中，由于异常的模式往往难以预测，因此，人为归纳的异常无法很好地覆盖所有异常类型。当面对新的异常时，以往的方法可能无法准确识别。

针对实际工业应用中存在的异常样本少、标注成本高、以往无监督检测方法计算开销较大、师生网络模型由于结构相似性高导致过模仿映射问题，提出了一种基于非对称师生网络的工业图像无监督异常检测方法 NTSNet(Nonsymmetric Teacher-Student Network)。本文主要工作如下：

(1)为了解决过模仿映射问题，即在检测异常区域时学生网络容易模仿教师网络的输出，产生与教师网络相似的特征映射，影响异常检测性能，提出在学生网络中引入上下文 Transformer 结构(Contextual Transformer Block，CoTB)，通过为学生网络引入自注意力机制，增加师生网络之间的结构差异性，由此缓解师生网络之间存在的过模仿映射问题，提高检测性能。

(2)为了解决异常区域在师生网络之间的映射特征差异不足的问题，在教师网络中引入移动平均归一化层(Exponential Moving Average Normalization，EMAN)。教师网络中的 EMAN 层根据学生网络的 BN 层数据进行更新，减少 BN 层固有的交叉样本依赖性，增强师生网络之间的泛化性差异，从而进一步提高检测性能。

(3)由于异常区域大小不一，为了综合不同尺度之间的信息，提出使用多尺度加权融合策略，利用多尺度特征来检测不同大小的异常，提高对不同尺度异常的检测精度。

(4)在公开数据集 MVTecAD 上的多项实验验证了所提方法的有效性与可靠性。

2 相关工作

2.1 基于生成模型的异常检测方法

Bergmann 等[17]提出了 AE，通过自编码器重构输入图像，使模型拟合为只对正常图像很好的重构，而对异常图像的重构效果较差。Kingma 等[18]在自编码器的基础上提出 VAE，通过引入变分推断的思想以学习正常数据的潜在表示。Song 等[19]提出了 AnoSeg 异常检测网络，基于生成对抗网络，使用硬增强技术改变正态样本分布，生成合成异常图像和参考掩码。通过自监督学习，使用合成的异常数据和正常数据对 AnoSeg 使用逐像素和对抗性损失进行训练。Madan 等[20]提出了一种用于自编码器中即插即用的自监督掩蔽卷积变换器块，通过在自监督预测卷积注意块中引入三维掩盖卷积层、通道的注意机制和 Huber 损失，提高了检测性能。

2.2 基于嵌入向量的异常检测方法

Wan 等[21]提出了一种预训练特征映射框架，通过将图像从预训练的特征空间双向或者多级双向映射到另一个特征空间，以有效地检测异常。Cohen 等[22]提出了 SPADE，该方法通过预训练网络提取图像特征表示，并利用 KNN 算法对测试图像的特征进行分类。Gudovskiy 等[23]提出了 CFLOW，该方法基于条件归一化流框架，使用预训练编码器和多尺度生成解码器，估计编码特征的可能性。Salehi 等[24]提出了 MKD，该方法利用正常图像在教师网络的多尺度特征表示作为伪监督信息，指导学生网络学习正常图像表征。由于教师学生网络的泛化性不同，异常区域在两个网络中的映射输出存在差异，因而导致异常分数较高。Defard 等[25]提出使用预训练的卷积神经网络来进行图像的补丁嵌入，这些嵌入被用来建模多元高斯分布，从而获得正态类的概率表示。

在上述方法中，基于生成模型的方法存在重构效果差或者训练不稳定等问题。基于嵌入向量的方法如 SPADE 与 CFLOW 等方法虽然取得了不错的检测效果，但这些方法的计算开销较大，难以在实际工业应用中部署。而 MKD 方法具有较快的运行速度和较高的精度，但由于师生网络之间的高相似对称结构，存在着过模仿映射的问题，即在检测缺陷区域时学生网络容易模仿教师网络的输出，从而产生与教师网络相似的特征映射，影响缺陷检测性能。

3 非对称师生网络模型

3.1 网络模型总体结构

本文在基线方法 MKD 的基础上，提出了非对称师生网络模型 NTSNet，主要包括教师网络和学生网络两个部分。教师网络采用在大数据集 ImageNet 上预训练的 ResNet[26]。而为了解决过模仿映射问题，本文使用的学生网络在 ResNet 的基础上，将 ResNet 的残差块中用于提取特征的 3×3 卷积层替换为改进 CoTB 自注意力模块，通过引入与卷积层结构不同的 Transformer 模块，为师生网络增加网络的不对称性，从而缓解过模仿映射现象。

NTSNet 的具体结构如图 1 所示，左边为未训练的学生网络，右边为预训练的教师网络。通过冻结教师网络预训练

的参数层，利用教师网络指导学生网络学习，在训练和测试阶段，图像将同时输入教师网络和学生网络中。然而，整个训练过程中，只使用正常图像作为训练集，教师网络不进行反向传播，而学生网络则根据正常图像在教师网络的中间层特征输出以更新自身参数。通过最小化师生网络中间层映射之间的分布差异以训练 NTSNet 模型，使模型对正常区域产生较小的特征差异。在测试阶段，使用训练完成的学生网络和预训练教师网络进行推理。由于训练时只采用了正常图像，在面对正常区域时，教师网络和学生网络中间层输出之间的分布差异会比较小。而对于异常区域，由于异常区域的模式与正常区域相差较大，而且模型未学习异常的先验信息，加之教师网络和学生网络之间同时存在泛化性差异，因此在模型遇到异常区域时，相比于正常区域，师生网络的中间层输出则会产生较大的特征差异。利用这一原理，能够有效区分正常区域和异常区域。

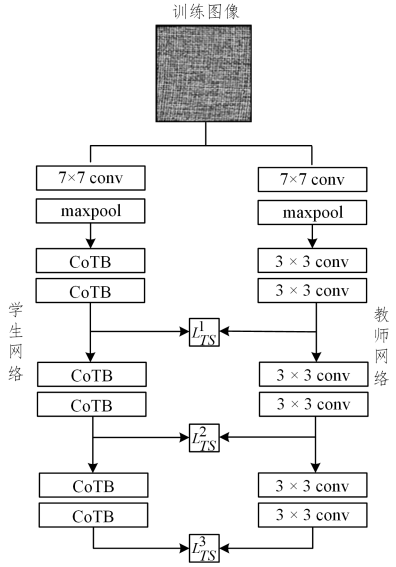

图 1　NTSNet 结构

Fig. 1　Nonsymmetric teacher-student network structure

3.2　改进上下文 Transformer 模块

CoTB[27]是一种基于 Transformer 的自注意力结构。它将 Transformer 模型捕获全局信息的能力与卷积神经网络对邻近局部信息的捕捉能力相结合，从而显著提升了网络模型的特征表达能力。

本文在 CoTB 模块基础上进行了改进，通过引入改进后的模块，可以增加师生网络的结构不对称性，从而缓解师生网络之间存在的过模仿映射问题，提高检测精度。改进后的 CoTB 模块的结构图如图 2 所示。

设输入特征图 $\boldsymbol{X}$ 的大小为 $C\times H\times W$，在生成值映射图 $\boldsymbol{V}$ 与映射图 $\boldsymbol{Q}$ 的过程中多增加了一个 1×1 的卷积层，以增加映射图的表征能力。输入特征 $\boldsymbol{X}$ 经过两个 1×1 的卷积后得到值映射图 $\boldsymbol{V}$，输入特征 $\boldsymbol{X}$ 经过一个 1×1 卷积后得到特征 $\boldsymbol{Q}$。同时，输入特征 $\boldsymbol{X}$ 经过 $k\times k$ 的卷积核进行卷积操作，实现邻近局部信息的捕捉，得到包含邻近间的上下文信息的局部特征 $\boldsymbol{K}$。然后，将局部特征 $\boldsymbol{K}$ 与特征 $\boldsymbol{Q}$ 进行叠加(*Concat*)操作，经过两个 1×1 卷积后，再进行 Softmax 层，之后与值映射图 $\boldsymbol{V}$ 进行自注意力计算，以获得图像全局信息特征 $\boldsymbol{G}$：

$$\boldsymbol{G}=\boldsymbol{V}\circledast\mathrm{Softmax}([\boldsymbol{K}+\boldsymbol{Q}]\boldsymbol{C}_\theta\boldsymbol{C}_\delta) \tag{1}$$

其中，$\boldsymbol{C}_\theta$ 与 $\boldsymbol{C}_\delta$ 表示 1×1 卷积操作，$\circledast$ 表示矩阵乘法计算。

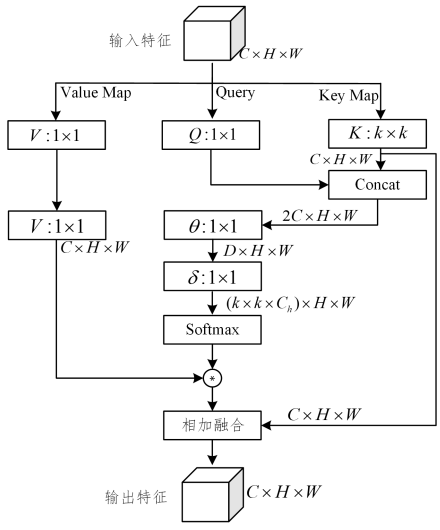

图 2　改进的上下文 Transformer 模块结构

Fig. 2　Improved contextual Transformer block structure

最后，将所得的邻近信息特征 $\boldsymbol{K}$ 与全局信息特征 $\boldsymbol{G}$ 进行相加融合，获得动态上下文输出特征 $\boldsymbol{X}'$：

$$\boldsymbol{X}'=\boldsymbol{K}+\boldsymbol{G} \tag{2}$$

由于动态上下文输出特征 X' 结合了邻近局部信息和全局上下文信息，这使网络能够更有效地理解输入图像的语义特征，从而提高学生网络的表征能力。而引入改进 CoTB 模块带来的非对称结构，则可以解决 MKD 高相似性结构带来的学生网络过模仿映射问题，从而增强师生网络对异常区域的映射差异，提高异常检测的性能。

3.3　移动平均归一化层

在传统的批量归一化(BN)方法中，数据的归一化是在每个批量中计算的，这容易导致模型过度依赖同一批数据中样本之间的相关性。实际上，BN 只是利用了数据批次内的一些统计性质，而未真正理解任务的本质，尤其在异常检测等任务中，样本之间的依赖性导致师生网络之间的泛化性差异减小，从而导致检测性能不佳。

为了解决这一问题，本文在教师网络中引入了 EMAN 层[28]来优化师生网络的性能。与标准 BN 不同，EMAN 层通过引入指数移动平均来更新数据，从而在一定程度上削弱了教师模型跨样本依赖性，增强了教师和学生网络之间的潜在泛化性差异。EMAN 层不仅能更加灵活地适应学生网络的 BN 层参数，而且通过整个训练过程中的均值和方差更新，减少了对当前批次数据的依赖，有助于提升模型对任务本质的理解。

值得注意的是，学生网络仍然使用 BN 层，这是为了保持对学生模型训练的有效性。EMAN 层只在训练阶段和测试阶段中被使用，在预训练阶段并未添加。通过在教师和学生网络中选择适用的归一化层，从而提升整体异常检测性能。

BN 层能够稳定学习，使收敛速度加快，因此被广泛采用。在训练过程中，BN 首先计算当前批处理数据 $\{\boldsymbol{x}_i\}_{i=1}^n$ 的均值与方差：

$$\mu_{\mathrm{BN}}=\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i \tag{3}$$

$$\sigma_{\mathrm{BN}}^2=\frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{x}_i-\mu_{\mathrm{BN}})^2 \tag{4}$$

其中，n 是批量大小，μ_{BN} 为均值，σ^2_{BN} 为方差。

接下来，使用分批统计的 μ_{BN} 和 σ^2_{BN} 对当前批次中的每个样本 $\boldsymbol{x}$ 进行标准化，然后应用具有可学习参数 γ 和 β 的仿射变换：

$$\hat{\boldsymbol{x}}=BN(\boldsymbol{x})=\gamma\frac{\boldsymbol{x}-\mu_{\mathrm{BN}}}{\sqrt{\sigma^2_{\mathrm{BN}}+\epsilon}}+\beta \tag{5}$$

而教师网络中的 EMAN 通过学生网络的 BN 层得到的均值 μ_{BN} 与方差 σ^2_{BN} 来更新参数，如下式所示：

$$\mu_{\mathrm{EMAN}}=m\mu_{\mathrm{EMAN}}+(1-m)\mu_{\mathrm{BN}} \tag{6}$$

$$\sigma^2_{\mathrm{EMAN}}=m\sigma^2_{\mathrm{EMAN}}+(1-m)\sigma^2_{\mathrm{BN}} \tag{7}$$

$$\hat{\boldsymbol{x}}=EMAN(x)=\gamma\frac{\boldsymbol{x}-\mu_{\mathrm{EMAN}}}{\sqrt{\sigma^2_{\mathrm{EMAN}}+\epsilon}}+\beta \tag{8}$$

其中，m 为接近 1 的常数，本文取 0.99；ϵ 是一个极小的常数，用于数值稳定性。

3.4 损失函数改进

在训练阶段，只有正常样本参与训练，经预处理后的训练样本被分别送入教师网络和学生网络中训练。对于单个输入图像 $\boldsymbol{X}_k\in\boldsymbol{R}^{w*h*c}$，其中 h 表示图像高度，w 表示图像宽度，c 表示颜色通道的数量。通过教师网络 $TNN(\cdot;\theta_t)$ 与学生网络 $SNN(\cdot;\theta_s)$，可以分别得到图像 $\boldsymbol{X}_k$ 在位置 (i,j) 的中间层映射特征 $\boldsymbol{F}^l_T(\boldsymbol{X}_k)_{ij}\in\boldsymbol{R}^{w_l*h_l*c_l}$ 与 $\boldsymbol{F}^l_S(\boldsymbol{X}_k)_{ij}\in\boldsymbol{R}^{w_l*h_l*c_l}$，即：

$$\begin{cases}\boldsymbol{F}^l_T(\boldsymbol{X}_k)_{ij}=TNN(\boldsymbol{X}_k;\theta_t)\\ \boldsymbol{F}^l_S(\boldsymbol{X}_k)_{ij}=SNN(\boldsymbol{X}_k;\theta_s)\end{cases} \tag{9}$$

其中，θ_t 和 θ_s 为神经网络中的参数，l 代表神经网络中的第 l 个 layer 层。由此，图像 $\boldsymbol{X}_k$ 在教师网络与学生网络的不同尺度特征被提取。

基线方法 MKD 损失函数采用了均方差和余弦相似度加权融合作为损失函数，但是对于本文提出的 NTSNet 网络结构，采用这种复杂的损失函数会使模型更容易受到不同损失项的影响，模型难以学到有效的特征表示。所以，为了实现 NTSNet 中教师学生网络之间的知识传递，本文采用更为简单的均方差损失作为特征之间的差异度量函数，这是因为均方差损失能有效地衡量两个特征之间的差异，并在训练过程中引导网络学习更好的特征表示。通过采用简化的损失函数，可以更好地适应 NTSNet 结构的非对称性，提高模型的训练稳定性和性能。对于输入图像的像素点 (i,j)，本文将教师网络 T 与学生网络 S 之间第 l 层的损失函数 $\mathcal{L}_{\mathrm{TS}}$ 的公式描述如下：

$$\mathcal{L}^l_{\mathrm{TS}}=\frac{1}{2w_lh_l}\sum_{i=1}^{w_l}\sum_{j=1}^{h_l}\left\|\frac{\boldsymbol{F}^l_T(\boldsymbol{X}_k)_{ij}}{\|\boldsymbol{F}^l_T(\boldsymbol{X}_k)_{ij}\|_2}-\frac{\boldsymbol{F}^l_S(\boldsymbol{X}_k)_{ij}}{\|\boldsymbol{F}^l_S(\boldsymbol{X}_k)_{ij}\|_2}\right\|^2_2 \tag{10}$$

其中，w_l 和 h_l 表示映射特征的高度和宽度，$\|\cdot\|_2$ 表示均方差距离。

将每个残差层后的输出作为嵌入特征，计算师生网络不同层之间的损失，最后的损失函数如式所示：

$$\mathcal{L}_{\mathrm{total}}=\sum_{l=1}^{L}\alpha_l\mathcal{L}^l_{\mathrm{TS}} \tag{11}$$

其中，L 是所使用特征层的数量；α_l 是为每个层次赋予的权重，为了简便运算，本文取 $\alpha_l=1$。

3.5 多尺度加权融合

测试与推理阶段算法流程如图 3 所示，在此阶段，由于训练时只采用了正常图像，因此相比于异常区域，正常区域在师生网络之间的映射差异会更小，而异常区域的特征映射差异会更为显著，这使得在特征空间中形成了具有区分性的映射差异。首先通过 NTSNet 得到师生网络的中间层特征，由于使用了多个中间层进行特征映射，因此通过式(10)计算第 l 层特征的异常分数，可以获得不同尺度的异常分数图 M^l。

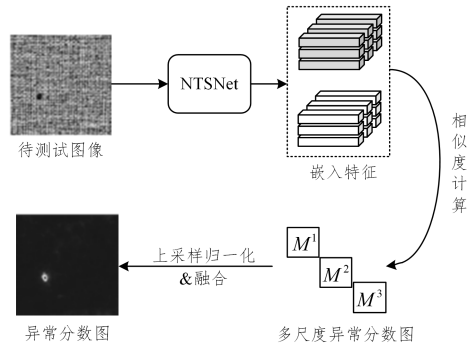

图 3 测试阶段算法流程图

Fig. 3 Algorithm flow chart of test phase

由于异常区域的大小不一，为了综合不同尺度的信息，以检测不同尺度大小的异常，提高检测性能，本文将多个尺度的异常分数图进行融合。由于不同层次之间的异常分数图的尺度不同，每个异常图 M^l 首先通过双线性插值被上采样至相同的大小，然后采用高斯平滑使异常图的异常边界更加清晰，最后将其所有上采样后的异常图相乘融合，得到整个图像对应的最终异常图 $\boldsymbol{M}_{\mathrm{final}}$：

$$\boldsymbol{M}_{\mathrm{final}}=\sum Upsample(g_\sigma(\boldsymbol{M}^l)) \tag{12}$$

其中，g_σ 为高斯平滑滤波器，σ 表示其参数；$Upsample$ 表示双线性上采样。

取异常图 $\boldsymbol{M}_{\mathrm{final}}$ 中所有像素对应的最大异常分数为单个图像最后的异常分数，即：

$$S=\max(\boldsymbol{S}_{ij}),\boldsymbol{S}_{ij}\in\boldsymbol{M}_{\mathrm{final}} \tag{13}$$

由此，每张测试图像都会得到一个异常分数。由于正常图像和异常图像的异常分数会有差异性，因此可以利用这一差异性区分正常图像和异常图像。

4 实验

4.1 实验环境设置

本文的实验环境为 Windows 11 操作系统，搭载 Intel Core i5-12600K CPU，RTX 3070TiGPU，配备 8 GB 显卡内存和 32 GB 系统内存。使用 Adam 算法优化，初始学习率为 0.4，weight decay 为 0.0001，训练 Epochs 为 200，批次大小设置为 32，输入图像大小为 256×256，采用了 ResNet 作为主干网络，深度学习框架为 PyTorch。

4.2 实验数据集与指标介绍

本文的实验选择在 MVTec AD 公共数据集[29]上进行训练、验证和测试，该数据集来自 MVTec 公司，用于对侧重于工业图像的异常检测方法进行测试。MVTecAD 共有 5 000 多张高分辨率图像，分为 15 种不同的对象和纹理类别。其中训练集有 3629 张图像，全部由无标注的正常图像组成，测试集有 1 725 张图像，由正常图像和异常图像混合组成，其中异常图像均做了像素级标注，用于验证模型的检测性能。训练时只采用训练集中的正常图像，测试时使用混合正常图像和异常图像的测试集测试性能，以衡量模型对正常图像和异常图像的区分能力。

实验部分使用 ROC 曲线下的面积(AUROC)作为衡量

指标，包括图像级别和像素级别的 AUROC。AUROC 是最广泛使用的异常检测评价指标之一。对于 AUROC 指标来说，较高的值表示模型具有更好的检测能力（最高值为 1）。ROC 曲线的横坐标为假阳率（FPR），纵坐标为真阳率（TPR），其中 FPR 为正常图像被错误分类为异常图像的百分比，TPR 为异常图像被正确分类的百分比。

$$TPR=\frac{TP}{TP+FN} \tag{14}$$

$$FPR=\frac{FP}{FP+TN} \tag{15}$$

其中，TP 为正常图像被正确分类的数量，FN 为正常图像被错误分类的数量，FP 为异常图像被错误分类的数量，TN 为异常图像被正确分类的数量。

4.3 实验结果与分析

4.3.1 与其他先进方法的对比实验

本文将提出的方法与长期存在的经典检测方法 AE[17]、VAE[18]、最近提出的先进的的方法 SPADE[22]、CFLOW[23]，以及基线方法 MKD[24]，经过复现后进行性能比较，实验结果如表 1 所列。

表 1 MVTec AD 数据集对比实验结果

Table 1 Comparative experimental results of MVTec AD datasets

（%）

类别	AE-L2[17]	VAE[18]	SPADE[22]	CFLOW[23]	MKD[24]	本文方法
Carpet	65.8,87.3	67.5,73.5	92.8,97.2	96.7,98.6	95.4,**98.6**	**97.8**,98.3
Grid	85.8,94.2	83.5,96.1	47.3,93.7	96.1,96.8	98.2,**98.8**	95.2,98.7
Leather	56.4,78.5	71.0,92.5	95.4,97.6	99.3,98.3	98.9,99.1	**100**,**99.5**
Tile	48.8,59.3	81.4,65.4	**96.5**,87.4	94.3,96.8	94.9,94.6	95.9,**97.9**
Wood	61.8,73.1	89.1,83.8	95.8,88.5	95.7,92.4	96.1,94.9	**98.8**,**96.9**
Bottle	83.2,93.5	86.8,92.2	97.2,98.4	96.8,98.1	97.9,97.1	**100**,**98.7**
Cable	48.6,82.3	57.0,91.5	84.8,**97.5**	**93.5**,95.5	83.8,89.8	91.3,96.8
Capsule	86.3,94.0	86.0,91.7	89.7,99.0	93.4,98.8	85.9,96.2	**93.7**,98.5
Hazelnut	92.1,97.3	74.3,97.6	88.1,99.1	96.6,**98.5**	**99.9**,98.1	99.8,98.4
Metal nut	60.4,89.5	78.3,90.7	71.0,98.1	91.6,**98.2**	95.6,94.2	**98.6**,96.7
Pill	84.0,91.2	80.1,93.3	80.1,96.5	**95.3**,97.3	80.5,87.8	**93.7**,96.8
Screw	89.0,96.0	71.8,94.5	66.7,**98.9**	**95.3**,97.9	83.5,98.3	92.1,98.5
Toothbrush	78.4,92.3	89.3,98.5	88.9,97.9	95.0,98.5	**99.7**,98.3	91.4,**98.9**
Transistor	73.1,90.7	70.1,92.0	90.3,**94.1**	81.4,88.7	**95.3**,83.8	91.6,87.5
Zipper	67.3,88.5	67.3,86.9	96.6,96.5	**96.6**,98.0	94.5,97.2	96.3,**98.5**
平均值	72.1,87.2	77.0,89.3	85.4,96.0	94.5,96.8	93.3,95.1	**95.7**,**97.4**

注：左边为图像级别的 AUROC 分数，右边为像素级别的 AUROC 分数，加粗的数据为取得最好效果的数据。

从表 1 中可以看出，本文提出的 NTSNet 在图像级别和像素级别的 AUROC 指标中的平均值上取得了最好的性能，比基准 MKD 方法的图像级 AUROC 高出 2.4%，比像素级别的 AUROC 高出 2.3%，相比于 CFLOW 方法同样高出 1.2% 和 0.6%，其中对象类的 Bottle 与纹理类的 Leather 在图像级别的 AUROC 上取得了 100% 的性能，即将所有图像正确分类。这说明所提出的 NTSNet 模型很好地缓解了师生网络之间存在的过模仿映射问题，提高了异常检测精度，验证了所提出方法的适用性和有效性。

为了更加直观地对比所提出方法与其他方法的检测性能差异，本文将不同算法的异常检测定位可视化图进行了对比，实验结果如图 4 所示，其中每一列代表不同方法的异常定位效果图。可以看出，相比其他方法，本文提出的方法展现出较好的异常定位效果。

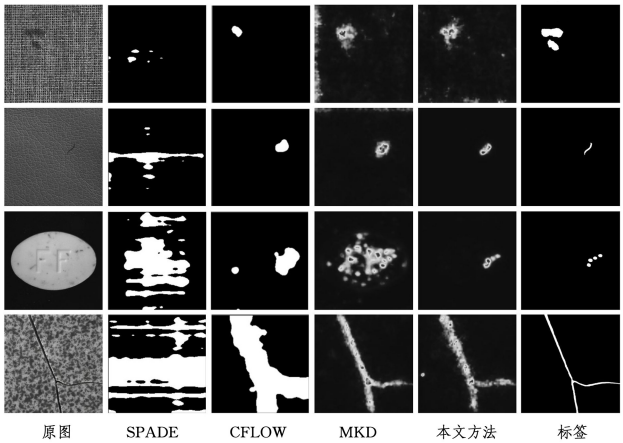

图 4 异常区域定位效果对比图

Fig. 4 Comparison of anomaly region localization effects

另外，为了更好地评估所提出方法的有效性，本文在检测速度和模型参数量方面与其他方法进行了对比实验，实验结果如表 2 所列，其中 FPS 表示每秒处理图像的数量，数值越大代表检测速度越快。相比于其他方法，本文所提出的方法在参数量上只需 17.3 MB，与其他方法相比有着参数量少的优势。在检测速度方面，本文方法排名第二，达到 6.8 FPS；SPADE 方法的检测速率只有 0.21FPS，在对比方法中最慢；MKD 方法在所有方法中的检测速度最快。与 MKD 方法相比，在没有损失太多检测速度的情况下，本文提出的方法在检测精度上有着不错的提升，这表明所提出的方法在模型轻量化和高效检测方面取得了显著的优势。

表 2 检测速度与参数量对比

Table 2 Comparison of detection speed and parameter number

方法	检测速度/FPS	参数量/MB
SPADE	0.21	170
CFLOW	3.5	96
MKD	**7.8**	22.4
本文算法	6.8	**17.3**

4.3.2 消融实验

为了评估模块的有效性，本文在 MVTecAD 数据集上对每个模块进行了消融实验，通过在 baseline 方法 MKD 上逐个添加模块的方式，对每个模块进行独立评估，并比较性能指标，以此综合评价模块的有效性，不同模块的消融实验结果如表 3 所列。从表 3 的实验结果可以看出，使用各个模块指标相较于基准方法都有所提高，其中单独使用改进 CoTB 模块的提升巨大，相比基准方法 MKD 在图像级别 AUROC 上提升了 1.7%，在像素级别的 AUROC 上提升了 1.8%。同时使用改进 CoTB 模块和 EMAN 模块的效果最好，只使用 EMAN

模块相比原方法也取得性能上的改善。这些消融实验表明了采用的每个模块对模型检测性能都有着一定的提升。

表 3 不同模块的消融实验结果

Table 3 Ablation results of different modules

模型	图像级别 AUROC/%	像素级别 AUROC/%
MKD	93.3	95.1
MKD_{+CoTB}	95.0	96.9
MKD_{+EMAN}	93.5	95.5
$MKD_{+CoTB+EMAN}$	**95.7**	**97.4**

4.3.3 主干网络与中间层层数选取研究实验

为了深入研究主干网络与中间层层数对检测性能的影响，本文选取了不同规格的主干网络和不同中间层层数进行了实验。由于中间层输出的特征大小要保持一致，所以本文采用了 ResNet18 和 ResNet34 作为主干网络。主干网络选取实验如表 4 所列，每一行上方代表教师主干网络的选取，下方代表学生主干网络的选取，教师网络采用预训练参数，学生网络采用随机初始化的参数，教师网络不参与反向传播，学生网络参数不断更新。

表 4 不同主干网络选取下的实验结果

Table 4 Experimental results of different backbone network selection

教师网络	学生网络	图像级别 AUROC/%	像素级别 AUROC/%
ResNet18	ResNet18	95.7	97.4
ResNet34	ResNet34	95.2	96.4
ResNet18	ResNet34	94.5	96.1
ResNet34	ResNet18	95.5	97.0

从表 4 中可以得知，教师网络和学生网络都采用 ResNet18 作为主干网络时效果最好。使用特征提取能力更强的网络并没有取得更好的效果，这是因为在训练样本数量较少的情况下，较小的网络更容易拟合数据，过大的网络容易导致过拟合现象产生，无法很好地学习正常样本的信息，从而影响检测性能。

中间层层数选取实验结果如表 5 所列。本文选择了 ResNet 网络中 3 个不同中间层次的映射输出特征进行实验，它们分别来自 ResNet 中 layer1，layer2，layer3 残差块后的输出特征。从表 5 可以看出，只使用单个层次特征的检测效果较差，随着使用层数的增加，模型的检测效果逐渐提高，其中使用 3 个层次的效果最好，这是因为利用多个层次的输出特征能够综合不同尺度的信息，从而使模型能更全面地理解和捕捉图像特征。

表 5 不同中间层选取下的实验结果

Table 5 Experimental results under different intermediate layer selections

使用层数	图像级别 AUROC/%	像素级别 AUROC/%
只使用 layer1 层	93.6	93.1
只使用 layer2 层	94.1	93.3
只使用 layer3 层	93.5	94.2
使用 layer1，2 层	95.5	96.8
使用 layer1，3 层	94.3	96.1
使用 layer2，3 层	95.1	96.9
使用 layer1，2，3 层	**95.7**	**97.4**

图 5 给出了所提出方法的微小缺陷检测定位可视化图。从图中可以看出，通过将不同尺度得到的异常分数进行融合，可以综合多个尺度的异常分数图信息，从而更加准确地定位异常区域。

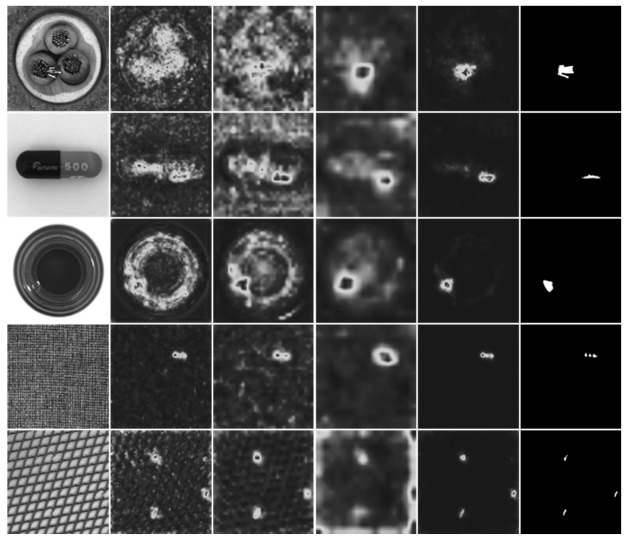

图 5 本文方法的图像异常检测定位可视化图

Fig. 5 Visualization of anomaly image detection and localization using the proposed method

结束语 针对工业图像异常检测问题，本文提出了一种基于非对称师生网络的工业图像无监督异常检测模型 NTS-Net。该方法通过使用改进 CoTB 模块替换 3×3 卷积层，缓解了师生网络之间存在的过模仿映射问题。此外，在教师网络中引入了移动平均归一化层，增强了师生网络之间的泛化性差异。最后，引入的多尺度异常图融合策略，通过综合不同尺度的异常图信息，提高了检测精度。本文所提出的方法在 MVTec AD 数据集上进行了对比实验，实验结果表明，所提出的方法有效地提升了检测精度。

虽然本文提出的非对称师生网络方法在提高精度方面取得了显著成果，但仍存在一些改进空间。可以考虑进一步改进非对称师生网络的结构，探索更多有效的注意力机制和特征融合策略，以提升异常检测的准确性和鲁棒性。

参 考 文 献

[1] 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.

[2] LOWE D G. Distinctive image features from scale-invariant keypoints[J]. International Journal of Computer Vision, 2004, 60(2): 91-110.

[3] ARRIS C, STEPHENS M. A combined corner and edge detector[C]// Alvey Vision Conference. 1988.

[4] WANG Z, WU F, HU Z. Msld: a robust descriptor for line matching[J]. Pattern Recognition, 2009, 42(5): 941-953.

[5] ZHANG L, KOCH R. An efficient and robust line segment matching approach based on lbd descriptor and pairwise geometric consistency[J]. Journal of Visual Communication and Image Representation, 2013, 24(7): 794-805.

[6] NISTÉR D, STEWÉNIUS H. Linear time maximally stable extremal regions[C]// Computer Vision-ECCV 2008: 10th European Conference on Computer Vision, Marseille, France, Part II 10. Springer Berlin Heidelberg, 2008: 183-196.

[7] XU Y, MONASSE P, GÉRAUD T, et al. Tree-based morse regions: a topological approach to local feature detection[J]. IEEE

Transactions on Image Processing,2014,23(12):5612-5625.
[8] GIRSHICK R,DONAHUE J,DARRELL T,et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2014:580-587.
[9] CZIMMERMANN T,CIUTI G,MILAZZO M,et al. Visual-based defect detection and classification approaches for industrial applications—A survey[J]. Sensors,2020,20(5):1459.
[10] REN S,HE K,GIRSHICK R,et al. Faster r-cnn:towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence,2017,39(6):1137-1149.
[11] YANG L,ZHONG J,ZHANG Y,et al. An improving faster-rcnn with multi-attention resnet for small target detection in intelligent autonomous transport with 6g [J]. IEEE Transactions on Intelligent Transportation Systems,2023,24(7):7717-7725.
[12] REDMON J,DIVVALA S,GIRSHICK R,et al. You only look once:unified,real-time object detection[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016:779-788.
[13] JIANG P,ERGU D,LIU F,et al. A review of yolo algorithm developments[J]. Procedia Computer Science, 2022, 199: 1066-1073.
[14] LIU W,ANGUELOV D,ERHAN D,et al. Ssd:single shot multibox detector[C]// Computer Vision-ECCV 2016:14th European Conference, Amsterdam, The Netherlands, Part I 14. Springer International Publishing,2016:21-37.
[15] ZHENG W,TANG W,JIANG L,et al. Se-ssd:self-ensembling single-stage object detector from point cloud[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021:14494-14503.
[16] SAMARIYA D,THAKKAR A. A comprehensive survey of anomaly detection algorithms[J]. Annals of Data Science,2023,10(3):829-850.
[17] BERGMANN P,LWE S,FAUSER M,et al. Improving unsupervised defect segmentation by applying structural similarity to autoencoders[C]// 14th International Conference on Computer Vision Theory and Applications. 2019.
[18] POL A A,BERGER V,GERMAIN C,et al. Anomaly detection with conditional variational autoencoders[C]// 2019 18th IEEE International Conference on Machine Learning And applications (ICMLA). IEEE,2019:1651-1657.
[19] SONG J,KONG K,PARK Y I,et al. AnoSeg:anomaly segmentation network using self-supervised learning [EB/OL]. (2021-10-7) [2024-02-19]. https://doi. org/10. 48550/arXiv. 2110. 03396,2021.
[20] MADAN N,RISTEA N C,IONESCU R T,et al. Self-supervised masked convolutional transformer block for anomaly detection [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2023,46(1):525-542.
[21] WAN Q,GAO L,LI X,et al. Unsupervised image anomaly detection and segmentation based on pretrained feature mapping [J]. IEEE Transactions on Industrial Informatics,2022,19(3):2330-2339.
[22] COHEN N,HOSHEN Y. Sub-image anomaly detection with deep pyramid correspondences [EB/OL]. (2020-05-05)[2024-02-19]. https://doi. org/10. 48550/arXiv. 2005. 02357,2020.
[23] GUDOVSKIY D,ISHIZAKA S,KOZUKA K. Cflow-ad:real-time unsupervised anomaly detection with localization via conditional normalizing flows[C]// Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). 2022:98-107.
[24] 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.
[25] DEFARD T,SETKOV A,LOESCH A,et al. Padim:a patch distribution modeling framework for anomaly detection and localization[C]// International Conference on Pattern Recognition. Cham:Springer International Publishing,2021:475-489.
[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] LI Y,YAO T,PAN Y,et al. Contextual transformer networks for visual recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2022,45(2):1489-1500.
[28] CAI Z,RAVICHANDRAN A,MAJI S,et al. Exponential moving average normalization for self-supervised and semi-supervised learning[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021:194-203.
[29] 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.

KONG Senlin, born in 1997, master. His main research interests include unsupervised learning and industrial image defect detection.

ZHANG Hui, born in 1983, Ph. D, professor, Ph. D supervisor. His main research interests include image processing and robot vision detection.