

中图法分类号: TP391.41 文献标识码: A 文章编号: 1006-8961(2025)08-2675-15

论文引用格式: Feng J, Meng X J, Shang Y Q and Niu C F. 2025. Adapter-enhanced two-stage continual defect detection. Journal of Image and Graphics, 30(8): 2675-2689 (封筠, 孟旭静, 尚玉全, 牛超凡. 2025. 结合适配器增强的双阶段连续缺陷判别. 中国图象图形学报, 30(8): 2675-2689) [DOI: 10.11834/jig.240663]

结合适配器增强的双阶段连续缺陷判别

封筠^{1,2*}, 孟旭静^{1,2}, 尚玉全^{1,2}, 牛超凡³

1. 石家庄铁道大学信息科学与技术学院, 石家庄 050043; 2. 石家庄市人工智能重点实验室, 石家庄 050043;
3. 河北圣昊光电科技有限公司, 石家庄 050000

摘要: 目的 传统异常检测方法在工业产品缺陷判别中仅关注当前任务, 从而导致在接受新任务训练时会灾难性地遗忘以前学过的知识。鉴于现实工业场景中对异常检测模型的灵活性和持续适应性的需求, 结合连续学习方法提出一种适配器增强的双阶段连续缺陷判别方法(adapter-enhanced two-stage continual defect detection, AETS)以实现连续异常检测任务。**方法** 1)在AdaptFormer基础上引入外部注意力机制, 增强模型对顺序任务中的全局依赖关系的捕捉能力, 以提升对新任务的泛化性能。2)在视觉转换器(vision Transformer, ViT)预训练模型的基础上结合高效微调技术, 采用双阶段训练策略, 即在适应阶段, 通过全量微调缓解自然图像与工业图像之间的域差异; 在高效微调阶段, 通过适配器增强模块提升模型对新任务的适应性, 同时冻结大部分参数以保留对旧任务的记忆, 从而缓解灾难性遗忘问题。3)提出遗忘波动率(forgetting fluctuation rate, FFR)这一新的连续学习评价指标, 用于量化模型在整个学习过程中遗忘波动情况, 以检验模型在工业场景中的适用性和稳定性。**结果** 在MVTec-MCIL、MVTec-SCIL和MVTec+MTD数据集上进行实验, AETS的ACC(average accuracy)值分别达到84.21%、89.16%和78.49%, 相较于5种连续学习方法, AETS具有最佳的ACC、FM(forgetting measure)值和最小的训练参数量, 相较于6种先进高效微调方法, 其FFR值达到最佳。消融实验选取缩放因子及确定适配器增强模块结构, 以实现模型可塑性与稳定性的最佳平衡。**结论** 所提出的AETS方法通过构建适配器增强模块, 充分利用预训练模型的特征表达能力, 双阶段训练策略能够捕捉与任务相关的特征, 显著增强模型在连续工业缺陷判别任务中的适应性和泛化性。**关键词:** 连续学习(CL); 工业产品缺陷判别; 异常检测; 适配器增强; 双阶段训练; 参数高效调整(PET); 遗忘波动率(FFR)

Adapter-enhanced two-stage continual defect detection

Feng Jun^{1,2*}, Meng Xujing^{1,2}, Shang Yuquan^{1,2}, Niu Chaofan³

1. School of Information Science and Technology, Shijiazhuang Tiedao University, Shijiazhuang 050043, China;
2. Shijiazhuang Key Laboratory of Artificial Intelligence, Shijiazhuang 050043, China;
3. Hebei Tahoo Photoelectric Technology Co., Ltd., Shijiazhuang 050000, China

Abstract: Objective The difficulty in obtaining defect samples in industrial scenarios causes many existing anomaly detection algorithms to rely solely on normal samples during the training phase. These methods demonstrate strong performance in industrial defect detection by training a single model tailored to specific product types. However, traditional anomaly detection methods mainly focus on the current task in industrial product defect discrimination, which frequently results in

收稿日期: 2024-11-11; 修回日期: 2025-01-23; 预印本日期: 2025-01-30

* 通信作者: 封筠 fengjun@stdu.edu.cn

基金项目: 国家自然科学基金项目(61972267)

Supported by: National Natural Science Foundation of China (61972267)

catastrophic forgetting knowledge of previously learned tasks when the model is trained on new tasks. As product objects and defect types continuously change and diversify in real-world industrial environments, anomaly detection models need to possess greater flexibility and continuous adaptation capability to cope with new tasks and data. Therefore, this study proposes an adapter-enhanced two-stage continual defect detection (AETS) method tailored for continuous anomaly detection tasks. **Method** First, based on the AdaptFormer framework, this study introduces an external attention mechanism to augment the capability of the model to capture global dependencies across sequential tasks, which enhances generalization performance on new tasks. This enhancement is crucial in industrial anomaly detection, which enables the AETS model to adapt to diverse and complex defect types across different products. Furthermore, combining parameter-efficient tuning technique with vision Transformer (ViT) pre-trained models, the AETS framework incorporates two primary stages of training to enhance the ability of the model to retain previously learned knowledge while effectively adapting to new tasks. The first stage focuses on adaptation, where a full parameter fine-tuning strategy is employed to mitigate the domain shift between natural and industrial images. This procedure is critical given that the differences in feature distributions between the two domains can severely hinder the performance of industrial anomaly detection models when directly applied to industrial datasets. The AETS model integrates an external attention mechanism into the existing AdaptFormer architecture to further enhance the adaptation process. This integration allows the model to capture long-range dependencies across sequential tasks, which is vital for maintaining performance across different domains and tasks. This incorporation of external attention enables the model to better generalize to new tasks by improving the representation learning of global contextual information, which not only improves the ability of the model to attend to relevant features but also enhances its capacity to model global contextual information. By refining the representation learning of the model, the external attention mechanism improves the flexibility and robustness. In the second stage, which is known as efficient fine-tuning, the AETS method utilizes adapter modules to selectively fine-tune a minimal number of parameters. This way reduces computational overhead while retaining essential knowledge from previously learned tasks. By freezing most of the parameters, the model can preserve knowledge from prior tasks and mitigate catastrophic forgetting, which is a common challenge in continual learning settings. The adapter-enhanced module is designed to enhance the capacity of the model to handle new tasks without overwriting previously learned information. This capability ensures that the model can continuously adapt to evolving data distributions in industrial environments. Forgetting measure (FM) only evaluates the forgetting compared with the maximum performance achieved in the past for each task at the completion of the final model training. This study introduces a novel evaluation metric called forgetting fluctuation rate (FFR), which is specifically designed for continuous learning scenarios to provide a comprehensive assessment of the catastrophic forgetting problem. FFR is used for quantifying the extent of forgetting during the learning process and can provide a more granular assessment of model stability across sequential tasks. By measuring the fluctuation in forgetting over time, FFR helps evaluate how well the model retains its knowledge while learning new tasks, which offers a more robust evaluation of continual learning performance in industrial defect detection scenarios. When the FFR value is low, the forgetting ability of the model is relatively smooth, with less performance fluctuation. Conversely, higher values mean greater fluctuations in forgetting. **Result** We conduct extensive experiments using three benchmark datasets from the MVTecAD dataset, which are specifically divided to evaluate the performance of the model in different industrial defect detection scenarios. The AETS method demonstrates significant improvements in average accuracy (ACC). Specifically, it achieves scores of 84.21%, 89.16%, and 78.49% on the MVTec-MCIL, MVTec-SCIL, and MVTec+MTD datasets, respectively. These results indicate that AETS outperforms other continual learning models, especially in terms of efficiency and resource utilization. Compared with five continuous learning methods, AETS exhibits the best performance in terms of ACC and FM metrics, with the smallest number of training parameters. Furthermore, when compared with six state-of-the-art parameter-efficient tuning methods, AETS achieves optimal FFR values. This result demonstrates that our method not only improves model accuracy but also provides a lightweight solution that is more practical for real-world industrial applications. We perform ablation studies to further validate the effectiveness of our proposed method. Specifically, the ablation experiments focus on selecting the optimal scaling factor and determining the best structure for the adapter enhancement module. These experiments are crucial in achieving a balanced trade-off between the plasticity of the model, which enables it to adapt to new tasks, and its stability, which helps

preserve previously learned knowledge. **Conclusion** This study presents a novel continual anomaly detection method called AETS, which leverages the powerful feature representation capability of pre-trained models combined with adapter-enhanced modules to improve continual learning performance in industrial defect detection tasks. AETS, which adopts a two-stage training process, effectively addresses domain shift issues between natural and industrial images. It also efficiently captures task-relevant features through parameter-efficient fine-tuning, which not only mitigates catastrophic forgetting but also enhances the training efficiency of the model. In addition, the introduction of FFR validates the robustness and adaptability of the model in dynamic industrial environments. The experimental results demonstrate that AETS can achieve competitive performance across various benchmark datasets. It not only maintains high detection accuracy but also significantly enhances the stability and plasticity of the model in the context of continuous learning. By mitigating catastrophic forgetting, AETS effectively enables the model to adapt to new tasks without compromising previously learned knowledge. These advantages make AETS particularly well suited for real-world industrial applications, where accuracy and the ability to continuously learn from evolving data are essential for practical deployment and long-term performance.

Key words: continual learning (CL); industrial product defect detection; anomaly detection; adapter enhancement; two-stage training; parameter-efficient tuning (PET); forgetting fluctuation rate (FFR)

0 引言

在现代工业制造过程中,光学检测广泛应用于产品质量控制,但一个常见的问题是缺陷样本往往不足,尤其是针对未知或罕见缺陷类型的检测。这使得基于监督学习的模型在缺乏充足标注样本时,难以准确识别异常。因此,许多现有的异常检测算法(王文鹏等,2024)依赖于无监督方法定位可能的异常区域。深度学习方法在工业缺陷检测中展示出强大的性能(白艳峰等,2024;王素琴等,2024),尤其是在针对特定产品或固定数据集的情况下,训练单一模型通常能取得出色的检测效果。然而,这些模型大多基于封闭世界假设,即假定训练和测试的数据分布是独立同分布的。在实际的工业生产环境中,这一假设存在明显的局限性。生产线可能会引入新产品类型、出现不同批次的生产波动,甚至可能产生新的缺陷形式等(倪一鸣和陈松灿,2022)。传统的深度学习模型在面对动态变化的场景时,性能可能会显著下降。因此,如何在不断变化的环境中具备持续学习能力,保持对未知情况的敏感性,已成为工业缺陷检测领域的一个重要研究方向。

传统的深度学习模型在开放环境中的适应性较为有限,主要因为其缺乏应对数据动态变化的能力。而在现实世界中,数据通常以流的形式不断产生,伴随隐私保护和存储资源限制等问题,使得长时间存储这些流数据变得不切实际。因此,工业缺陷检测亟需一种能够持续适应新数据并累积知识的学习系

统,这一能力称为连续学习(continual learning, CL),其核心挑战在于如何有效平衡模型的可塑性与稳定性。可塑性决定模型在面对新任务和新数据时的泛化能力,而稳定性则确保模型不会遗忘对先前任务的记忆。如何在动态变化的环境中保持这种平衡,是确保连续学习成功应用的关键(Kirkpatrick等,2017;李文斌等,2024;周大蔚等,2023)。

在连续学习中,“从头开始训练”的无监督学习方法通常需要大量样本来捕捉鲜明的特征。然而,对于较小规模、单一类别的异常检测任务,这种方法往往难以提供足够的检测能力。近年来,连续学习从预训练模型(pre-trained model, PTM)(Han等,2021)的应用中受益显著。PTM强大的泛化能力可为连续学习提供有效的知识迁移,这对提高连续学习方法的性能至关重要(Mehta等,2023)。然而,当逐步适应顺序到达的新任务时,如果覆盖或过度修改预训练的知识,可能会削弱PTM的优势。因此,Zhou等人(2024)将参数高效调整(parameter-efficient tuning, PET)技术引入连续学习,通过冻结预训练模型的主干网络,仅引入少量轻量级参数进行表示学习,既能适应新任务,又能保留原有知识。

Li等人(2022)提出一种用于连续异常检测任务的正态嵌入分布(distribution of normal embeddings, DNE)方法,采用PTM来执行图像级的异常检测任务,通过在训练过程中固定分类头参数而仅微调特征提取器中的参数,并结合特征重放缓解灾难性遗忘问题。然而,这种全量微调除具有较高的计算开销外,还可能在任务间引入灾难性干扰,影响模型的

稳定性。针对这些问题,本文提出一种基于适配器增强的双阶段连续缺陷判别方法(adapter-enhanced two-stage continual anomaly detection, AETS),旨在将大型预训练的视觉转换器(vision Transformer, ViT)模型有效迁移到工业领域的下游任务中。这一方法通过仅微调少量参数,能够保留预训练模型知识,同时显著提升迁移学习能力,可以避免任务间的灾难性干扰。

本文的贡献包括:1)提出一种双阶段连续学习训练框架,用于缓解灾难性遗忘。在适应阶段,对预训练模型进行参数微调,以缩小PTM模型与工业领域知识的差距;在高效微调阶段,仅调整少量参数,既能适应新任务又能保持对旧任务的记忆。2)改进AdaptFormer的适配器模块,通过将其与外部注意力机制(external attention, EA)并联连接,以增强适配器模块在高效微调阶段的特征提取能力,从而提升模型在连续学习过程中的泛化性。3)提出遗忘波动率(forgetting fluctuation rate, FFR)这一新的评价指标,通过量化模型在整个顺序学习过程中的遗忘波动性,更加精确地评估模型的稳定性,进一步验证连续学习方法在复杂工业场景中的适用性和鲁棒性。

1 相关工作

1.1 连续学习与异常检测

经典的连续学习方法主要分为动态架构方法、经验重放方法和正则化方法。动态架构方法通常为每个新任务分配子网络结构或参数空间,避免学习新任务时对旧任务知识产生干扰,从而实现低遗忘。一种典型的做法是冻结模型的某些参数,增加模型容量来学习新任务。经验重放方法通常将部分先前任务的样本保留在记忆空间(memory bank, MB)中,或通过生成模型获得伪样本。由于样本存储涉及隐私问题,并且占用大量内存,因此一些方法使用中间层特征进行样本回放。记忆空间或生成的样本在学习新任务时被重放,以保持从先前任务中学习的信息,其做法简单且有效。正则化方法通常设计模型参数的更新过程,而不是依赖先前任务样本的记忆空间。大多数正则化策略的目标是提高学习权重或其激活的稳定性。一些常用的方法包括惩罚重要权重的更新或通过知识蒸馏进行正则化。

在将连续学习策略应用于工业缺陷检测领域的

以往研究中,大多集中于单一场景或特定对象上。这些方法通过逐步引入新的缺陷类型,有助于模型更好地捕捉和区分不同类型的缺陷,从而提升模型在特定场景下的检测性能。Yildiz等人(2022)引入连续学习判别晶体线缺陷,根据缺陷分类器的准确性动态生成训练和推理数据,可显著提高分类器精度,同时减少数据样本的使用。Chang等人(2022)提出一种增量YOLO(you only look once)架构实现轮胎气泡缺陷检测,旨在当遇到训练数据集中不存在的新缺陷类别时,能够增量训练模型以识别该类缺陷。Dekhovich和Bessa(2025)提出一种连续学习分割带钢和磁瓦缺陷算法,通过迭代剪枝方式为每个缺陷类型创建对应子网络,并基于线性判别分析预测缺陷类型。然而,当面对多样化的产品对象和复杂的生产线时,这类方法在真实工业环境中的适用性依然有限。Li等人(2022)将连续学习方法应用于一般工业场景,提出一个解决连续异常检测(continual anomaly detection, CAD)问题的框架,通过无监督特征嵌入的异常检测,针对任务序列进行连续学习,并存储正常样本的特征分布以应对异常检测任务。Bugarin等人(2024)全面分析连续学习框架下的像素级异常检测问题,采用重放技术探讨更适合连续学习的异常检测方法,为该领域的进一步研究奠定了基础。

1.2 预训练模型的参数高效调整

先前的连续学习研究大多集中在“从头开始”训练模型,但随着研究的深入,越来越多的工作开始探索PTM的技术优势(江静和邓伟洪,2020;Zhou等,2024)。PTM不仅提供强大的知识迁移能力,还能够缓解灾难性遗忘问题,使其在下游任务中表现出色(Mehta等,2023)。然而,当PTM用于顺序到达的任务时,如何有效提取预训练知识并确保持续适应性,仍是需要面临的挑战。频繁微调可能导致预训练知识的覆盖,而单独保存每次微调的模型又会带来额外的资源负担(Zhang等,2023)。

为解决在连续学习过程中全量微调模型带来的上述问题,本文提出的AETS方法采用PET技术,通过调整少量附加参数,使预训练模型适应下游任务。PET与全量微调相比在成本上更具优势,且能够实现与全量微调相媲美甚至更优的性能。例如,视觉提示调整(visual prompt tuning, VPT)方法(Jia等,2022)通过在输入层或隐藏层添加轻量级的可训练

提示模块,实现预训练模型的高效微调。Chen等人(2022)提出一种并行适配器结构 AdaptFormer,利用缩小和放大投影提升视觉分类的增量调整性能。多认知视觉适配器(multi-cognitive visual adapter, Mona)方法(Yin等,2023)则通过引入多个过滤器,并在适配器中添加比例归一化层,优化视觉信号处理能力和输入特征的分布。低秩适应(low-rank adaptation, LoRA)方法(Hu等,2021)通过学习低秩矩阵来近似参数更新,以减少参数调整的计算成本。缩放和移动深层特征(scale and shift the deep features, SSF)方法(Lian等,2022)通过缩放和移位操作有效地解决模型调整中的性能问题。此外,Bahng等人(2022)还提出在输入空间中学习可调参数的方案,进一步提高连续学习的适应性。

基于预训练的无监督异常检测方法通常利用在大规模图像数据集上训练得到的权重参数初始化模型的特征提取器。通过进一步微调,这些预训练特征得以适应目标数据集的特定知识,其目的是使特征提取器能够更精准地捕捉目标数据的分布特性。Reiss等人(2021)研究表明结合预训练特征与基本异常检测方法可以进一步提升检测性能。然而,直接微调可能导致特征崩溃,从而降低模型的泛化能力。为克服这一问题,一些研究(Defard等,2021; Lee等,2018; Rippel等,2021)基于高斯先验的方法,通过将特征分布建模为多元高斯分布,并利用马氏距离评估异常分数,以此在保留预训练特征优势的同时,避免特征崩溃发生。本文基于上述思想提出一种融合预训练模型与高斯先验的连续学习框架,旨在实现对异常样本的高效检测,并增强模型在不断变化的工业环境中的适应能力。

2 本文工作

2.1 问题定义

无监督异常检测在训练阶段只对正常样本进行建模来判定待测样本是否存在异常。为描述该任务,将多类数据集定义为 $D = \{D^1, D^2, \dots, D^n, \dots, D^N\}$, 其中, $D^n = \{(x_i, y_i)\}_{i=1}^{m_n}$, $y \in \{0, 1\}$ 表示第 n 类别的样本, N 表示多类工业产品数据集的产品类别数。显而易见的是,不同工业产品类别的数据分布是不同的,但是具有相同的输出标签空间。

本文重点关注无范例的连续学习设置,即在无法获取历史样本进行重放训练的情况下,逐步为所有可见的子任务构建一个统一的模型。具体而言,多类数据集被划分为 T 个任务 $\{D_t\}_{t=1}^T$, 其中每个任务 t 包含 c 类产品,而且不同任务间的类集合是不相交的。模型在子训练集 D_t 上进行顺序训练,并在所有过去的子测试集 $D_t^{\text{total}} = \{D_1, D_2, \dots, D_t\}$ 上评估模型的性能。这种评估方法可确保经 t 个任务训练的模型能够保留先前获得知识的能力。

2.2 模型架构

受适应和合并(adapt and merge, ADAM)方法(Zhou等,2024)的启发,为减少预训练模型与工业领域下游任务之间的域偏差,本文采用双阶段优化策略:在第1阶段即适应阶段,对预训练模型进行全量微调参数,以适应工业领域的特定知识,将微调后的模型作为特征提取器。通过将多元高斯分布拟合到提取的深度特征正常数据上,并使用马氏距离作为异常评分机制,可以有效识别潜在的异常样本。在第2阶段即高效微调阶段,为适应新任务的同时保留模型对旧任务知识的记忆,以减少模型可训练参数并防止模型过拟合,仅对预训练模型的 ViT 模块进行部分微调参数,即完全冻结图像补丁嵌入层(patch embedding, PE)以及 ViT-EA 块中的层归一化(layer norm, LN)、多头自注意力(multi-head self-attention, MHSA)和多层感知机(multi-layer perceptron, MLP)参数,仅对适配器增强模块进行训练。为缓解模型的灾难性遗忘,并考虑到样本的隐私和存储效率问题,本文采用高效的特征重放技术,将每个任务的特征分布存储到记忆空间中。这种两阶段的训练策略不仅可提高模型对新任务的适应能力,而且可实现有效的知识迁移,从而缓解连续学习模型的灾难性遗忘问题。本文 AETS 的结构如图1所示。

在训练阶段,采用在大规模数据集上预训练的 ViT 模型初始化网络,其中, $\phi(\cdot)$ 表示特征提取器, h 表示分类头。将输入图像 x 投影到嵌入空间 $z = \phi(x) \in \mathbf{R}^{B \times L}$, 其中 B 代表批次大小, L 为嵌入向量的维度。与 GaussianAD(Rippel等,2021)相同,假设这些嵌入特征 $z_1, \dots, z_n \in \mathbf{R}^L$ 服从高斯分布,其均值 μ 和协方差矩阵 $\hat{C}$ 能够描述数据的分布特性,计算式为

$$\mu = \frac{1}{n} \sum_{i=1}^n z_i \quad (1)$$

$$\hat{C} = \frac{1}{n-1} \sum_{i=1}^n (z_i - \mu)(z_i - \mu)^T \quad (2)$$

在通常情况下极大似然估计用来近似真实的协方差矩阵,但这要求观测值的数量 n 远大于特征维数 L 。然而,在实际的异常检测数据集中,样本数量有限且深度特征维度较高,这可能导致经验协方差矩阵的估计不稳定。为解决此问题,本文采用收缩(shrinkage)方法(Ledoit 和 Wolf, 2004)估计协方差,通过减少经验协方差矩阵的最大和最小特征值之间的比率,增强协方差矩阵的稳定性,计算式为

$$C = (1 - \alpha)\hat{C} + \alpha \frac{\text{tr}(\hat{C})}{L} I_d \quad (3)$$

式中, I_d 为单位矩阵, α 表示收缩权值(亦称收缩强度), $\alpha \in [0, 1]$ 。如果参数 α 的值为 0,则返回样本矩阵,表示没有收缩;如果参数 α 的值为 1,则表示存在完全收缩。

随着训练时任务的变化,模型的权重可能会与当前任务数据发生过拟合。这是因为存储在记忆空间中的先前任务对应的特征分布,不再与新模型的输出特征相匹配。本文提出一种优化的训练策略以解决该问题,首先在第 1 个任务上对特征提取器进行全量微调,以使模型适应工业领域的特定知识。在后续的任务学习过程中,仅对适配器增强模块进

行微调,而保持其他参数不变。其目的是有效抑制特征分布中的过度偏差,从而减少对先前任务知识的遗忘,提升模型在多任务场景下的稳定性和泛化能力。

在推理阶段,利用记忆空间中保存的知识帮助模型进行预测。具体而言,基于记忆空间中的统计数据 (μ_t, C_t) ,其中 $t = 1, 2, \dots, T$,分别为每个旧任务构建对应的高斯分布。然后,从这些分布中抽取 n 个样本点,其代表新的高斯分布样本,这样将构建一个包含所有旧任务特征的总体高斯分布 $\varphi_t(\mu, C)$,即使在未知测试图像来自何种具体任务的情况下,该模型也能通过特征分布推断图像是否异常。最后,采用马氏距离作为异常评分的标准,马氏距离用于度量特定样本与均值 μ 和协方差 C 所定义的高斯分布之间的距离,其定义为

$$M(x) = \sqrt{(x - \mu)^T C^{-1} (x - \mu)} \quad (4)$$

2.3 适配器增强模块

为应对顺序任务数据分布的变化而导致的灾难性遗忘问题,传统的全量微调方法往往容易导致模型在旧任务上的表现大幅下降。为进一步增强连续学习模型的泛化能力,本文基于 AdaptFormer(Chen 等, 2022)提出一种引入外部注意力的即插即用式适配器增强模块,旨在鼓励知识的正向迁移,弥补不同

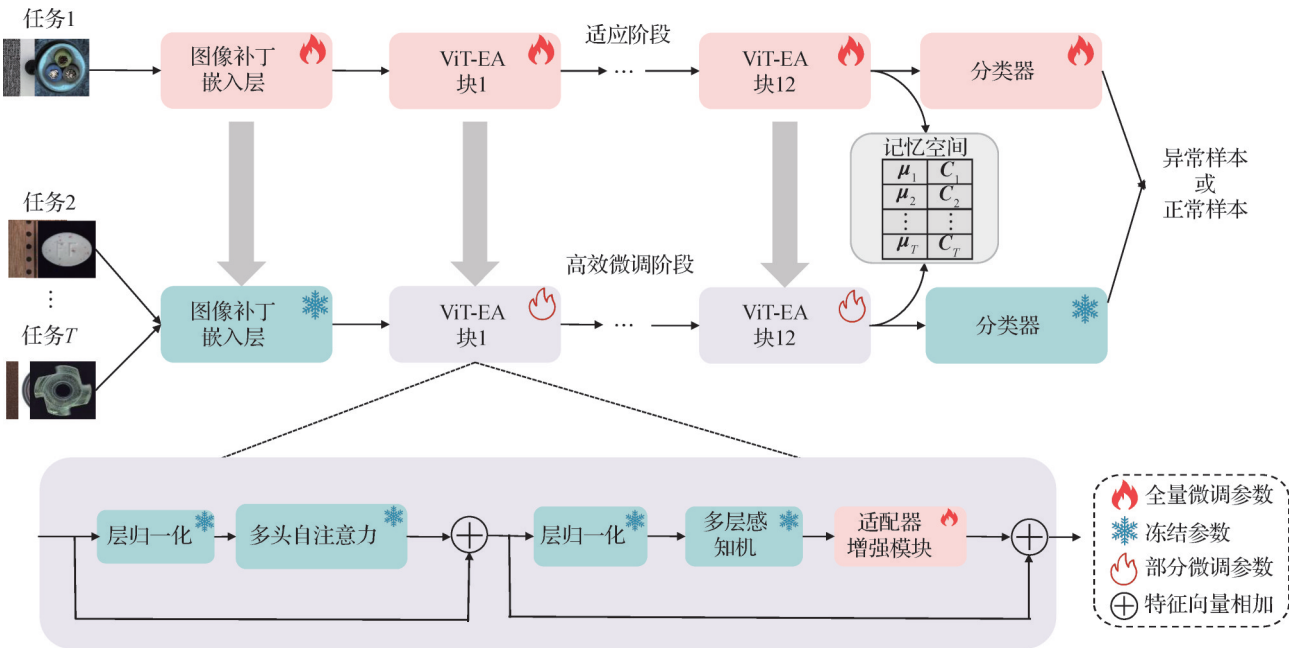


图1 AETS整体结构示意图

Fig. 1 Schematic diagram of the overall structure of AETS

任务间的领域差距。

与传统的全量微调方法相比,AdaptFormer采用瓶颈结构设计,包括1个用于降维的向下投影层 $\mathbf{W}_{\text{down}} \in \mathbf{R}^{d \times r}$ 和1个用于恢复原始维度的向上投影层 $\mathbf{W}_{\text{up}} \in \mathbf{R}^{r \times d}$ 。在这两个投影层之间,还加入一个非线性激活函数 ReLU(rectified linear unit)层。这种结构设计可大大减少模块的参数数量,从而降低计算复杂度。

外部注意力机制(Guo等,2023)在捕捉序列数据中的全局上下文信息方面发挥着重要作用。与自注意力机制不同,外部注意力通过引入一个外部存储单元 $\mathbf{M} \in \mathbf{R}^{u \times d}$ 来增强模型对全局特征的学习能力。通过计算输入特征与外部存储单元之间的注意力权重,进而生成一组全局上下文表示,这些表示能够从数据集级别的先验知识中提取出重要信息,有效地弥合任务间的领域差距。与自注意力机制不同,外部注意力的计算独立于输入特征,因此计算成本较低,同时可捕捉跨任务的全局依赖关系。

本文提出的 AdaptFormer 适配器增强模块,旨在融入注意力优化模型的特征提取与表达能力。适配器及增强模块的结构如图2所示。MLP 与 AdaptFormer 的连接方式分为两种结构。图2(a)为串联连接,即将 MLP 的输出作为 AdaptFormer 的输入;图2(b)为并联连接,并行融合 MLP 与 AdaptFormer 的输出特征。图2(c)(d)分别在图2(a)(b)的基础上引入外部注意力 EA 模块,以进一步增强特征提取的灵活性和模型对不同任务的适应性。由3.3节的消融实验确定图2(c)为本文适配器增强模块的最优设计,其展示出面向复杂任务时显著的稳定性和可塑性。该适配器注意力增强模块由两个子分支组成,即 AdaptFormer 和外部注意力分别构成上、下分支。原始的 MLP 输出特征经过这两个分支的处理,AdaptFormer 负责对特征进行自适应增强,而 EA 模块则聚焦于外部上下文信息的整合。最终,通过缩放因子(scale)形成融合后的特征。对于 MLP 的输出特征 $\mathbf{x}_l$,由适配器注意力融合模块所产生的特征可以表述为

$$\mathbf{x}_l^{\text{out}} = s(f_{\text{ReLU}}(\mathbf{x}_l \mathbf{W}_{\text{down}}) \mathbf{W}_{\text{up}} + f_{\text{Norm}}(\mathbf{x}_l \mathbf{M}_k^T) \mathbf{M}_v) \quad (5)$$

式中, $\mathbf{W}_{\text{down}}$ 和 $\mathbf{W}_{\text{up}}$ 分别表示适配器的下投影层和上投影层,这些投影层之间还有一个 f_{ReLU} 层,用于实现非线性处理; $\mathbf{M}_k$ 和 $\mathbf{M}_v$ 分别表示两个不同的存储单

元作为键和值; f_{Norm} 表示标准化操作; s 表示缩放因子。

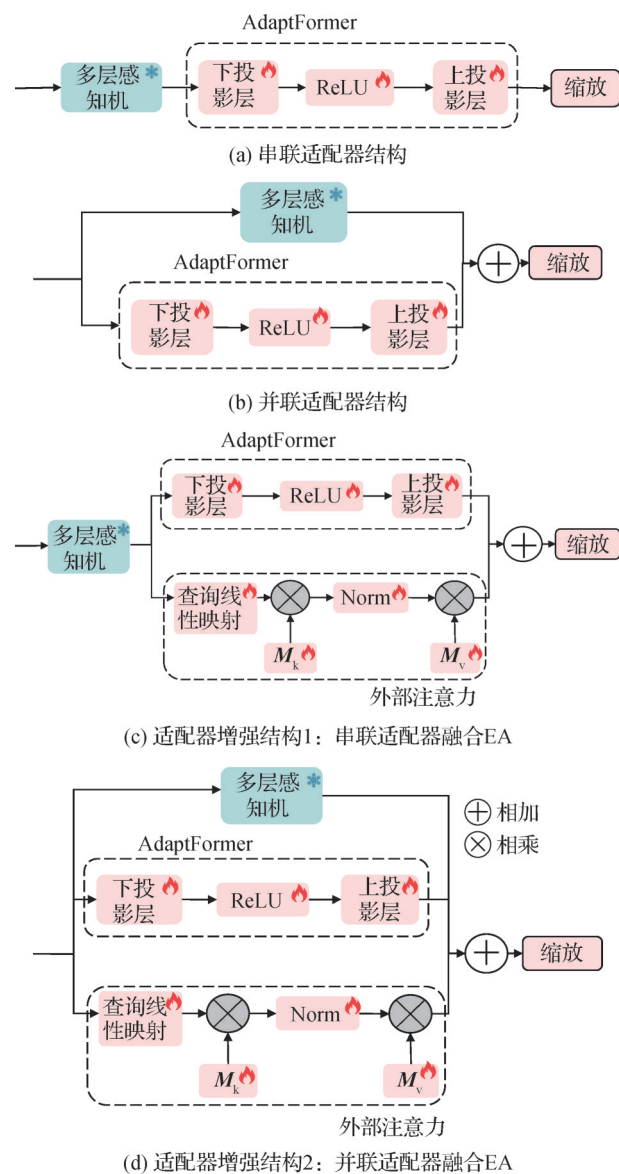


图2 适配器和适配器增强模块结构示意图

Fig. 2 Schematic diagrams of the structure of adapter and adapter enhancement module ((a) serial adapter structure; (b) parallel adapter structure; (c) enhanced adapter structure1: serial adapter fusion EA; (d) enhanced adapter structure2: parallel adapter fusion EA)

2.4 评价指标

为衡量模型在不同任务上的表现,连续学习常用的性能评价指标包括平均准确率(average accuracy, ACC)和平均遗忘率(forgetting measure, FM)。其中,ACC用于评估最终的分类模型在所有任务上的分类准确率,FM可以衡量连续学习方法对旧任务

的遗忘情况。定义任务 T 结束时任务 k 的准确率为 $R_{T,k}$, ACC 和 FM 分别表示为

$$ACC = \frac{1}{T} \sum_{k=1}^T R_{T,k} \quad (6)$$

$$FM = \frac{1}{T-1} \sum_{k=1}^{T-1} \max_{m \in \{1,2,\dots,T-1\}} (R_{m,k} - R_{T,k}) \quad (7)$$

灾难性遗忘是连续学习需要面对的关键问题之一。FM 作为一个评估此问题的重要指标,其主要关注模型在学习完所有任务后,相对于各个任务过去的最大性能的遗忘情况,存在如下的局限性:

1) FM 仅能够评估在最后一个任务训练结束时的遗忘情况,而没有反映模型在每个任务训练结束时的遗忘过程。

2) FM 没有记录模型在每次新任务学习后的逐步遗忘情况,也无法提供每个任务训练结束时的详细遗忘信息。

针对上述 FM 指标评估的不足,本文为更全面地评估模型在顺序学习过程中的性能,提出遗忘波动率 FFR 指标,通过量化模型在顺序学习过程中性能的波动性,提供更加全面的评估方式,其不仅捕捉平均遗忘情况,还可以衡量模型适应新任务时性能变化的一致性,即用于评估模型在不显著牺牲旧任务性能的情况下学习新任务的能力。

图 3 是连续学习模型在训练过程中的准确率变化情况示意图。其中,每行表示在新任务学习完成后,得到的模型在之前任务上的测试结果。例如,第 1 行第 1 列代表任务 1 (T_1) 训练结束后在其测试数据上的准确率表示为 $R_{1,1}$; 第 5 行则表示在任务 5 (T_5) 训练结束后,模型在 T_1 到 T_5 上的测试结果表示为 $R_{5,1} \sim R_{5,5}$ 。

通过计算每个任务在学习后续任务时准确率变化的标准差得出 FFR,从而直观展示模型在连续学习过程中遗忘波动的情况。FFR 的计算式可以表示为

$$FFR = \text{std}(\Delta P_1(2), \dots, \Delta P_1(T), \dots, \Delta P_{T-1}(T)) \quad (8)$$

$$\Delta P_i(t) = \ln R_{t,i} - \ln R_{t-1,i} \quad (9)$$

式中, $\Delta P_i(t)$ 表示对于任务 i , 模型在第 t 次学习之后相对于之前学习结束时的性能变化量。当 FFR 值较低时,意味着模型的遗忘较为平稳,性能波动较小,从而表明模型在任务切换时能有效保持稳定性,在适应新知识的同时可避免对旧知识的显著干扰。而当 FFR 值较高时,则反映出模型在多个任务之间的

遗忘波动较大,可能存在较为严重的灾难性遗忘,这意味着模型对新旧任务的平衡处理不当,稳定性较差。

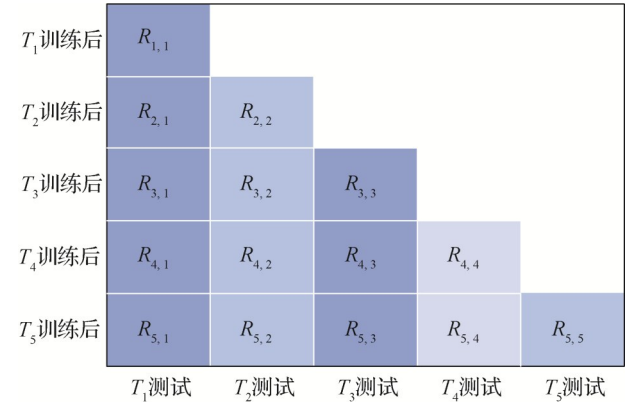


图 3 连续学习模型的准确率变化示意图

Fig. 3 The diagram of accuracy variation of continual learning model

3 实验及结果分析

3.1 实验数据集与任务划分

遵循 Rebuffi 等人 (2017) 提出的连续学习基准协议,即对于给定的多类分类任务数据集,以固定的随机顺序排列各个类。首先将所有类划分为 T 个任务,然后以增量方式在每个任务上训练模型,最后分类器在已经训练过的类别测试数据上进行评估,从而考察模型的连续学习能力。

MVTec AD (MVTec anomaly detection) (Bergmann 等, 2019) 是针对无监督异常检测的真实世界数据集,由 3 629 幅无缺陷训练图像和 1 725 幅异常图像组成,共包含 15 种产品(其中 5 种纹理类和 10 种物体类)。每种产品存在多种不同类型的缺陷。本文根据连续学习的任务设置,在 MVTec AD 基础上创建 3 个不同的数据集以模拟实际工业场景(Li 等, 2022),将 MVTec-MCIL、MVTec-SCIL 和 MVTec+MTD 统称为 seq-MVTec 数据集。

1) MVTec-MCIL。将 MVTec AD 划分为 5 个任务,其中每个任务包含 1 个纹理类和 2 个物体类产品。

2) MVTec-SCIL。将 MVTec AD 划分为 6 个任务,首先将 10 个产品类作为第 1 个任务,然后从第 2 个任务开始每个产品类为 1 个任务。

3)MVTec+MTD。为验证模型的泛化能力,引入另一个工业数据集,即MTD(magnetic tile defects)磁瓦表面缺陷数据集(Huang等,2020)。该数据集包含有缺陷和没有缺陷的磁瓦灰度图像,图像尺寸大小各异。与MVTec不同的是,MTD数据集中的无缺陷示例在照明条件和其他非缺陷特征方面存在较大的差异。为设定更具挑战性的场景,将MTD数据集作为第1个任务,并将MVTec-MCIL数据集的序列作为后续任务,共构成6个任务,进一步模拟复杂工业环境下的连续学习问题。

3.2 实验细节

数据增强策略可以通过生成伪样本,将异常检测从无监督任务转换为有监督任务,本文采用Cut-paste(Li等,2021)数据增强策略,将其应用于基于重放的有监督连续学习方法(如ER(experience replay)(Riemer等,2018)、DER(dark experience replay)(Buzzega等,2020)),以适用于异常检测场景。

实验硬件平台为Intel(R)Xeon(R)Silver 4214R CPU @ 2.40 GHz, NVIDIA GeForce RTX 4080, 编程语言为Python3.9, 深度学习框架为PyTorch。在模型训练阶段均使用Adam作为优化器,权重衰减系数为0.000 03。在第1阶段,使用ImageNet上训练得到的预训练权重,损失函数选择交叉熵损失函数,批量大小设为16。网络共训练50个轮次,初始学习率为0.000 1,采用余弦退火重启(cosine annealing warm restarts)作为学习率调度策略。在第1阶段训练结束后,冻结模型的大部分参数。第2阶段只训练适配器参数,初始学习率设为0.1,每学习一个任务降低学习率至当前的1/10,学习率最小为0.000 1。

将主流的高效微调方法(SSF、VPT、AdaptFormer、Mona、SSIAT(semantically-shifted incremental adapter-tuning))、微调(Fine-tuning)方法、样本重放方法(ER、DER)、特征重放方法(DNE)、模型扩展方法(memory-efficient expandable model, MEMO)、动态架构方法(ADAM)与本文的AETS方法进行对比实验。ER、DER依据原论文采用Resnet-18(residual network 18)作为backbone,记忆空间大小为200个样本。为将高效微调策略与连续学习框架相结合,SSF、VPT、AdaptFormer和Mona只对相应的网络结构进行修改。

3.3 实验结果

3.3.1 与现有文献方法的对比实验

为进一步验证本文所提出的适配器增强模块对缺陷判别性能的优越性,选择6种近年提出的先进高效微调方法进行对比。其中,SSF方法对预训练模型提取的深层特征进行缩放和平移,可以达到全量微调的性能;VPT方法将可调前缀标记添加到输入层或隐藏层,根据提示插入的位置,VPT可以进一步分为VPT-Deep和VPT-Shallow;Mona方法在适配器中引入多个滤波器,以增强其处理视觉信号的能力;SSIAT方法则增量微调共享适配器,且从存储的原型中进行特征采样来重新训练统一的分类器;AdaptFormer方法附加一个额外的适配器模块,在适应过程中模型冻结预训练的权重,仅优化额外的参数。此外,Upper Bound方法将所有的图像数据集合并训练,作为连续学习方法的参考上限,Fine-tuning方法表示全量微调。

表1为在seq-MVTec数据集上7种高效微调方法的对比实验结果。可以看出,在MVTec-MCIL和MVTec-SCIL上本文AETS方法的ACC、FM和FFR指标均取得最佳的结果,尤其适用于域不变的场景;在MVTec+MTD上AETS方法具有很强竞争力,其FFR达到最佳且ACC为次优。在6种对比的高效微调方法中,AdaptFormer方法的性能远远高于其他微调方法。与AdaptFormer方法相比,AETS方法在训练参数量不明显增长的前提下,在MVTec-MCIL和MVTec-SCIL上ACC值分别提高2.19%(84.21% vs 82.02%)和4.20%(89.16% vs 84.96%),FM值分别降低1.13%(1.65% vs 2.78%)和1.90%(0.94% vs 2.84%),FFR值分别降低0.64%(1.39% vs 2.03%)和0.84%(1.65% vs 2.49%);在MVTec+MTD上,AETS方法较AdaptFormer方法的FFR值降低1.20%(3.42% vs 2.22%),ACC值仅相差0.01%(78.49% vs 78.50%)。同时,AETS方法的性能均显著高于Fine-tuning方法,且更为接近参考上限Upper Bound方法,而AETS方法的训练参数量仅为这两种方法的2.70%。实验结果表明,与全量微调相比,仅微调部分参数不仅能够保持较高的准确率,而且可以显著提升模型在连续学习过程中的稳定性与泛化性。因此,相较于其他微调方法,本文提出的AETS方法不仅具有更高的准确率,还表现出更好的稳定性,即模型在顺序学习过程中波动率更低,能够在学习新任务知识的同时有效保持旧任务性能。

表 1 不同高效微调方法在 seq-MVTec 上的结果对比
Table 1 Comparison of results from different PET methods on the seq-MVTec

方法	训练参数量/M	MVTec-MCIL/%			MVTec-SCIL/%			MVTec+MTD/%		
		ACC ↑	FM ↓	FFR ↓	ACC ↑	FM ↓	FFR ↓	ACC ↑	FM ↓	FFR ↓
Upper Bound	87.58	88.07	1.94	2.44	91.70	0.47	1.86	84.65	0.89	1.17
Fine-tuning	87.58	68.53	25.69	14.90	77.06	9.66	21.11	73.49	18.39	25.11
SSF(Lian 等, 2022)	0.21	72.22	9.17	19.96	80.25	4.98	6.63	57.32	10.68	13.26
VPT-Shallow(Jia 等, 2022)	0.00	71.85	<u>1.89</u>	8.70	65.77	7.21	16.60	71.11	2.25	<u>3.10</u>
VPT-Deep(Jia 等, 2022)	0.05	58.38	13.52	20.91	61.64	14.96	19.74	60.25	13.19	21.41
Mona(Yin 等, 2023)	1.41	52.88	3.46	9.63	47.04	11.70	14.02	51.67	4.72	12.73
SSIAT(Tan 等, 2024)	1.20	47.75	8.17	18.29	45.65	4.47	5.83	57.01	15.61	14.25
AdaptFormer(Chen 等, 2022)	1.19	<u>82.02</u>	2.78	<u>2.03</u>	<u>84.96</u>	<u>2.84</u>	<u>2.49</u>	78.50	<u>2.32</u>	3.42
AETS(本文)	2.37	84.21	1.65	1.39	89.16	0.94	1.65	<u>78.49</u>	3.18	2.22

注:加粗、下划线字体表示各列最优、次优结果。“↑”表示值越大越好,“↓”表示值越小越好。

基于重放的方法是缓解模型在训练过程中的灾难性遗忘问题的简单有效方法。本文与 3 种重放连续学习方法进行比较。其中,ER 方法通过在新任务训练时,重放旧样本来缓解旧任务数据分布的遗忘;DER 方法结合重放、知识蒸馏和正则化,促进与其旧任务的一致性;DNE 方法采用特征重放方式,提高存储效率并强化隐私保护。本文提出的 AETS 方法是基于 DNE 方法进行改进的。此外,还与 MEMO、ADAM 方法进行对比,MEMO 是一种基于模型扩展的方法,通过共享浅层广义特征并为每个任务扩展专门的深层网络,以避免任务间的干扰;ADAM 充分利用 PTM 的泛化性,结合高效微调技术进一步提升 PTM 的适应性,通过融合 PTM 与自适应模型,保持连续学习模型的通用性。

表 2 为在 seq-MVTec 数据集上 6 种不同类型的连续学习方法的对比实验结果。可以看出,本文提出的 AETS 方法的训练参数量最小,仅为 DNE 和 ADAM 方法的 2.78% 和 1.66%,在 MVTec-MCIL 和 MVTec-SCIL 上 3 个性能指标均达到最优,在 MVTec+MTD 上 ACC、FM 达到最优且 FFR 为次优。与 DNE 方法相比,AETS 方法在 MVTec-MCIL、MVTec-SCIL 和 MVTec+MTD 上 ACC 值分别提高 2.17%(84.21% vs 82.04%)、6.91%(89.16% vs 82.25%)和 0.13%(78.49% vs 78.36%),FM 值分别降低 5.18%(1.65% vs 6.83%)、4.00%(0.94% vs 4.94%)和 6.01%(3.18% vs 9.19%),FFR 值分别降低 2.06%

(1.39% vs 3.45%)、1.19%(1.65% vs 2.84%)和 2.77%(2.22% vs 4.99%)。尤其是对于 FFR 值,本文提出的 AETS 方法在 MVTec-MCIL 和 MVTec-SCIL 上甚至比 Upper Bound 上界值分别降低 1.05%(1.39% vs 2.44%)和 0.21%(1.65% vs 1.86%)。另外,ADAM 方法虽然准确率不高,但其 FM 和 FFR 值比 ER 和 DER 方法都有明显降低,与 AETS 方法相比具有一定竞争力。由此可见,AETS 方法微调少量参数可以有效保留旧任务的知识,融合外部注意力的适配器增强模块能够学习到显著的判别特征,从而平衡模型学习新任务的可塑性和稳定性。因此,通过高效微调技术调整 PTM,能够有效提高持续学习模型的适应性,同时双阶段的训练策略可有效降低模型在顺序学习期间的波动情况,确保在应对新任务时不会过度干扰旧任务的性能,这对于需要长期部署和应用的连续学习模型尤为关键。

图 4 是对应于表 1 和表 2 中的 12 种不同方法在 seq-MVTec 的 3 个基准数据集上,分别经过对应类别学习后的逐任务 ACC 变化情况。在 MVTec-MCIL 和 MVTec-SCIL 上,本文 AETS 方法的逐任务 ACC 变化曲线均位于其他 11 种方法之上,尤其是在 MVTec-SCIL 上表现显著。与 DNE 方法的连续学习中间过程相比,AETS 方法也展现出极具竞争力的性能,其在适应阶段仅微调 2.78% 的参数量,却能达到与 DNE 方法全量微调接近的效果,显著降低了模型训练的计算成本,同时确保高效的任务适应能力。

表 2 不同类型的连续学习方法在 seq-MVTec 上的结果对比
Table 2 Comparison of results from different types of CL methods on the seq-MVTec

方法	训练参数量/M	MVTec-MCIL/%			MVTec-SCIL/%			MVTec+MTD/%		
		ACC ↑	FM ↓	FFR ↓	ACC ↑	FM ↓	FFR ↓	ACC ↑	FM ↓	FFR ↓
Upper Bound	87.58	88.07	1.94	2.44	91.70	0.47	1.86	84.65	0.89	1.17
Fine-tuning	87.58	68.53	25.69	14.90	77.06	9.66	21.11	73.49	18.39	25.11
ER(Riemmer等,2018)	11.77	59.15	20.71	15.40	52.39	12.31	12.17	52.79	19.60	13.57
DER(Buzzega等,2020)	11.77	67.37	12.62	14.11	64.63	7.57	15.77	56.74	16.85	9.84
DNE(Li等,2022)	85.21	<u>82.04</u>	6.83	<u>3.45</u>	<u>82.25</u>	4.94	<u>2.84</u>	<u>78.36</u>	9.19	4.99
MEMO(Zhou等,2023)	11.18	53.90	5.98	19.14	56.57	<u>1.27</u>	5.01	45.33	9.22	14.41
ADAM(Zhou等,2024)	143.09	51.77	<u>3.77</u>	4.33	50.09	1.70	2.85	51.15	<u>4.10</u>	1.97
AETS(本文)	2.37	84.21	1.65	1.39	89.16	0.94	1.65	78.49	3.18	<u>2.22</u>

注:加粗、下划线字体表示各列最优、次优结果。“↑”表示值越大越好,“↓”表示值越小越好。

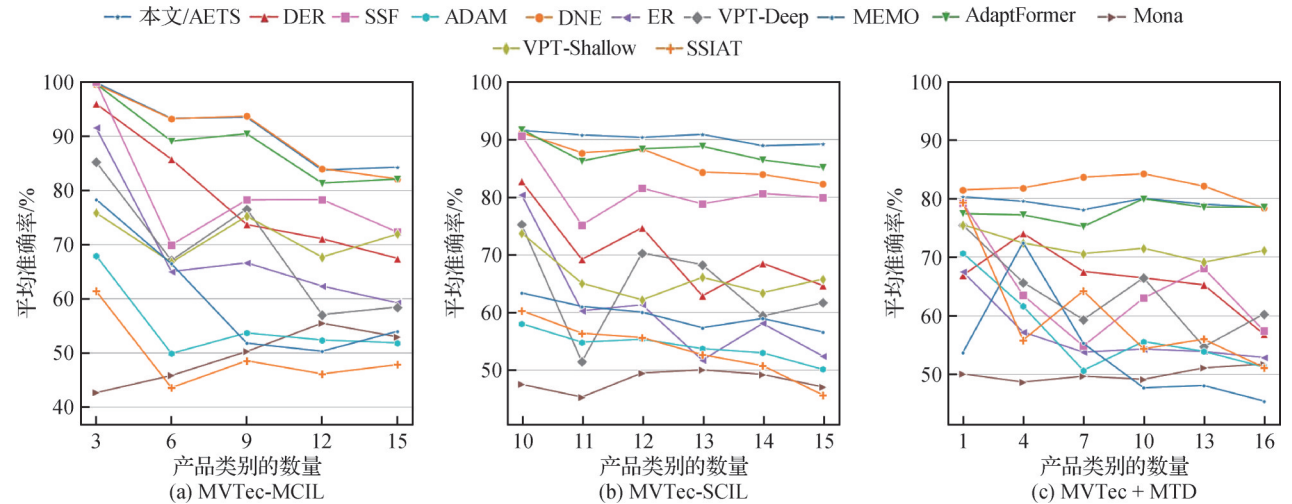


图 4 在 seq-MVTec 上 12 种方法逐任务的 ACC 变化

Fig. 4 Task-by-task ACC variation of twelve methods on seq-MVTec ((a) MVTec-MCIL; (b) MVTec-SCIL; (c) MVTec + MTD)

图 5 对比 DNE、ER、DER、AdaptFormer、ADAM 与本文 AETS 方法在 seq-MVTec 的 3 个基准数据集上,任务 1(T_1)的遗忘变化曲线。可以看出,在 MVTec-MCIL 和 MVTec-SCIL 上 AETS、AdaptFormer 方法相较于其他 4 种连续学习,其遗忘波动更为平稳。尽管 ADAM 方法的准确率明显低于其他方法,但是其遗忘变化曲线也较为平稳。由此可见,适配器模块在缓解模型灾难性遗忘方面效果显著,可有效保持模型在顺序学习中的稳定性。不过在 MVTec+MTD 上 AETS 和 AdaptFormer 方法的表现要逊色于 DNE 方法,可能的原因是适应阶段训练的参数偏少所导致。

3.3.2 适配器增强模块缩放因子消融实验

缩放因子 s 的作用在于平衡由原始冻结分支生成的与任务无关的特征和由适配器增强模块生成的特定任务的特征,其目的是在连续学习中折中模型对旧任务的记忆能力和对新任务的学习能力。通过消融实验探索不同缩放因子对模型性能的影响,表 3 给出在不同 s 值时的实验结果。当 $s = 0.10$ 时,在 seq-MVTec 的 3 个基准数据集上的 ACC 值均达到最优,相较于次优结果分别提升 0.54%(84.21% vs 83.67%)、0.69%(89.16% vs 88.47%)和 2.18%(78.49% vs 76.31%),FM 和 FFR 也表现出较好的效果。这表明缩放因子取 0.10 时可以很好地兼顾任务无关特征和特定任务特征,能够有效促进特征的

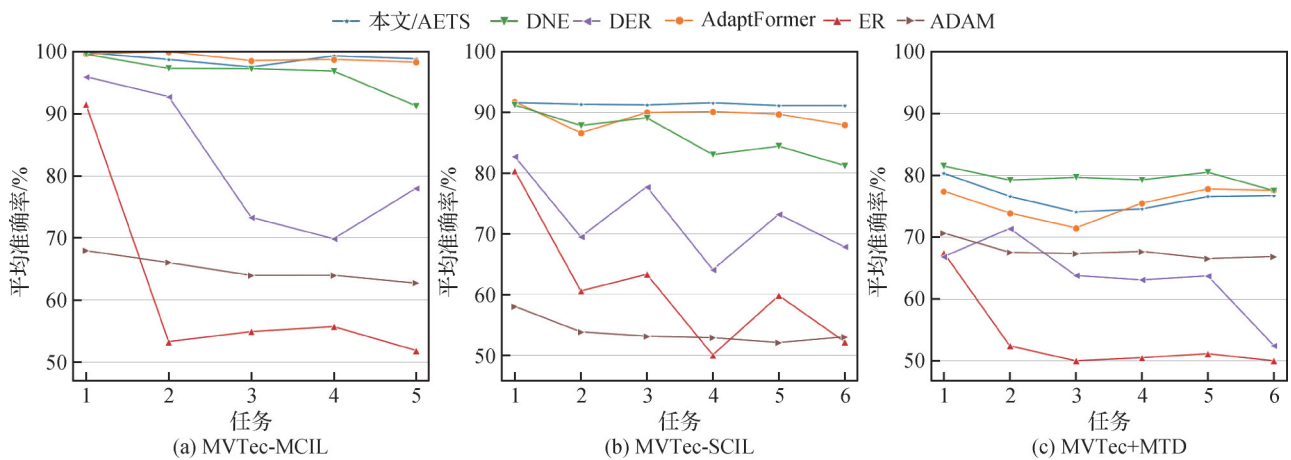


图5 在seq-MVTec上任务1的遗忘变化曲线

Fig. 5 The forgetting variation curve of the first task on seq-MVTec ((a) MVTec-MCIL; (b) MVTec-SCIL; (c) MVTec + MTD)

表3 在seq-MVTec上的缩放因子消融实验结果

Table 3 Ablation experiment results of the scaling factor on seq-MVTec

缩放因子 s	MVTec-MCIL/%			MVTec-SCIL/%			MVTec+MTD/%		
	ACC $\uparrow$	FM $\downarrow$	FFR $\downarrow$	ACC $\uparrow$	FM $\downarrow$	FFR $\downarrow$	ACC $\uparrow$	FM $\downarrow$	FFR $\downarrow$
0.01	81.67	2.85	4.35	<u>88.47</u>	<u>1.52</u>	<u>1.87</u>	76.04	<u>3.58</u>	2.20
0.05	<u>83.67</u>	1.27	<u>2.28</u>	86.74	3.52	4.15	75.93	4.10	3.02
0.10	84.21	<u>1.65</u>	1.39	89.16	0.94	1.65	78.49	3.18	<u>2.22</u>
0.30	81.97	3.82	2.97	83.64	3.80	6.84	<u>76.31</u>	4.40	3.02

注:加粗、下划线字体表示各列最优、次优结果。“ $\uparrow$ ”表示值越大越好,“ $\downarrow$ ”表示值越小越好。

互补,而偏小或偏大的缩放因子,因不能有效兼顾这两类特征,故不利于提升模型的分类准确率以及克服灾难性遗忘所带来的负面影响。因此,本文AETS方法最终设置 $s = 0.10$,以达到连续学习模型的可塑性和稳定性之间的更好平衡。

3.3.3 适配器增强模块结构消融实验

为验证本文提出的适配器增强模块结构的有效性,针对2.3节图2(a)—图2(d)这4种适配器模块结构设计进行消融分析,实验结果如表4所示。可以看出,模块结构(c)在MVTec-MCIL上获得最优的ACC、FM和FFR值,分别优于第2名1.46%(84.21% vs 82.75%)、1.61%(1.65% vs 3.26%)和0.78%(1.39% vs 2.17%);在MVTec-SCIL上获得最优的ACC和FM值,分别优于第2名0.85%(89.16% vs 88.31%)和0.40%(0.94% vs 1.34%);在MVTec+MTD上获得最优的FFR值,优于第2名0.31%(2.22% vs 2.53%)。

4种适配器模块结构的对比实验结果,进一步

揭示了将MLP输出特征作为适配器输入的顺序连接方式的有效性。MLP输出特征通过适配器增强结构中的两个分支分别处理特征的局部细节和全局上下文,可有效提高模型的适应性和表达能力。其中,AdaptFormer分支通过自适应机制优化特征的局部表达能力,可以根据具体任务的需求个性化地调整特征;而外部注意力分支则通过整合外部上下文信息来提升特征的全局关联性。最终,两个分支的输出通过缩放因子与冻结的多头注意力模块中的Norm层输出特征进行融合。冻结的多头注意力模块可提供一个稳定、通用的知识基础,有助于模型保持对已学任务的记忆,从而减轻在学习新任务时灾难性遗忘的发生。此外,表4中结构(a)与结构(c)的对比结果显示,模型不仅依赖于局部信息的自适应调整,还可以通过融入外部注意力机制整合全局上下文来增强特征的表达能力,特征的丰富性和多样性对于适应连续学习新任务至关重要。

表 4 在 seq-MVTec 上的适配器模块结构设计消融实验结果
Table 4 Ablation experiment results of adapter module structure design on seq-MVTec

模块结构	训练参数量/M	MVTec-MCIL/%			MVTec-SCIL/%			MVTec+MTD/%		
		ACC ↑	FM ↓	FFR ↓	ACC ↑	FM ↓	FFR ↓	ACC ↑	FM ↓	FFR ↓
(a)	1.19	<u>82.75</u>	<u>3.26</u>	2.38	85.10	3.04	2.53	75.34	3.20	4.31
(b)	1.19	79.03	4.99	<u>2.17</u>	83.77	2.70	3.02	78.76	3.79	2.77
(c)	2.37	84.21	1.65	1.39	89.16	0.94	<u>1.65</u>	<u>78.49</u>	<u>3.18</u>	2.22
(d)	2.37	78.10	5.08	4.48	<u>88.31</u>	<u>1.34</u>	1.05	76.02	2.26	<u>2.53</u>

注:加粗、下划线字体表示各列最优、次优结果。“↑”表示值越大越好,“↓”表示值越小越好。

4 结 论

本文提出一种结合适配器增强与双阶段训练策略的连续学习 AETS 方法,用于多个工业产品缺陷判别任务。该方法在 AdaptFormer 基础上引入外部注意力机制,构建适配器增强模块,并结合高效微调技术,采用双阶段训练策略,不仅能够捕捉与任务相关的特征,还可以避免模型在学习新任务时对已学任务产生过度干扰,从而在平衡模型稳定性与可塑性方面表现出显著优势。为全面评估连续学习方法的性能,本文提出遗忘波动率 FFR 指标,可以有效衡量模型在新旧任务上的一致性与稳定性。在 seq-MVTec 数据集上,通过消融实验选择适配器增强模块的结构和缩放因子,与近年先进文献方法的对比实验结果表明,本文提出的 AETS 方法在保持高检测精度的同时,可以显著提高模型在连续学习中的稳定性与可塑性,能够有效缓解灾难性遗忘问题。未来的研究将着重于进一步提升模型在应对跨领域任务时的泛化性,以及缓解任务顺序对模型遗忘程度的影响,从而适应更为复杂的工业场景和多样化的任务需求。

参考文献 (References)

Bahng H, Jahanian A, Sankaranarayanan S and Isola P. 2022. Visual prompting: modifying pixel space to adapt pre-trained models [EB/OL]. [2025-01-15].
<https://dblp.org/rec/journals/corr/abs-2203-17274.html>
Bai Y F, Wang L B, Gao W D and Ma Y L. 2024. Multi-modal hierarchical classification for power equipment defect detection. Journal of Image and Graphics, 29(7): 2011-2023 (白艳峰, 王立彪, 高卫东, 马应龙. 2024. 面向电力设备缺陷检测的多模态层次化

分类. 中国图象图形学报, 29(7): 2011-2023) [DOI: 10.11834/jig.230269]
Bergmann P, Fauser M, Sattlegger D and Steger C. 2019. MVTec AD—a comprehensive real-world dataset for unsupervised anomaly detection//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE: 9592-9600 [DOI: 10.1109/CVPR.2019.00982]
Bugarin N, Bugaric J, Barusco M, Pezze D D and Susto G A. 2024. Unveiling the anomalies in an ever-changing world: a benchmark for pixel-level anomaly detection in continual learning//Proceedings of 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 4065-4074 [DOI: 10.1109/CVPRW63382.2024.00410]
Buzzega P, Boschini M, Porrello A, Abati D and Calderara S. 2020. Dark experience for general continual learning: a strong, simple baseline//Proceedings of the 34th International Conference on Neural Information Processing Systems. Vancouve, Canada: Curran Associates Inc.: #1335 [DOI: 10.5555/3495724.3497059]
Chang C Y, Su Y D and Li W Y. 2022. Tire bubble defect detection using incremental learning. Applied Sciences, 12 (23): #12186 [DOI: 10.3390/app122312186]
Chen S F, Ge C J, Tong Z, Wang J L, Song Y B, Wang J and Luo P. 2022. AdaptFormer: adapting vision Transformers for scalable visual recognition//Proceedings of the 36th International Conference on Neural Information Processing Systems. New Orleans, USA: Curran Associates Inc.: #1212 [DOI: 10.5555/3600270.3601482]
Defard T, Setkov A, Loesch A and Audigier R. 2021. PaDiM: a patch distribution modeling framework for anomaly detection and localization//Pattern Recognition. ICPR International Workshops and Challenges: Virtual Event. Cham, Switzerland: Springer: 475-489 [DOI: 10.1007/978-3-030-68799-1_35]
Dekhovich A and Bessa M A. 2025. Continual learning for surface defect segmentation by subnetwork creation and selection. Journal of Intelligent Manufacturing, 36: 3051-3065 [DOI: 10.1007/s10845-024-02393-4]
Guo M H, Liu Z N, Mu T J and Hu S M. 2023. Beyond self-attention:

- external attention using two linear layers for visual tasks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(5): 5436-5447 [DOI: 10.1109/TPAMI.2022.3211006]
- Han X, Zhang Z Y, Ding N, Gu Y X, Liu X, Huo Y Q, Qiu J Z, Yao Y, Zhang A, Zhang L, Han W T, Huang M L, Jin Q, Lan Y Y, Liu Y, Liu Z Y, Lu Z W, Qiu X P, Song R H, Tang J, Wen J R, Yuan J H, Zhao W X and Zhu J. 2021. Pre-trained models: past, present and future. *AI Open*, 2: 225-250 [DOI: 10.1016/J.AIOOPEN.2021.08.002]
- Hu E J, Shen Y L, Wallis P, Allen-Zhu Z, Li Y Z, Wang S A and Chen W Z. 2021. LoRA: low-rank adaptation of large language models [EB/OL]. [2025-01-15].
<https://arxiv.org/pdf/2106.09685.pdf>
- Huang Y B, Qiu C Y and Yuan K. 2020. Surface defect saliency of magnetic tile. *The Visual Computer*, 36(1): 85-96 [DOI: 10.1007/S00371-018-1588-5]
- Jia M L, Tang L M, Chen B C, Cardie C, Belongie S, Hariharan B and Lim S N. 2022. Visual prompt tuning//*Proceedings of the 17th European Conference on Computer Vision*. Tel Aviv, Israel: Springer: 709-727 [DOI: 10.1007/978-3-031-19827-4_41]
- Jiang J and Deng W H. 2020. Facial expression recognition improved by continual learning. *Journal of Image and Graphics*, 25(11): 2361-2369 (江静, 邓伟洪. 2020. 持续学习改进的人脸表情识别. *中国图象图形学报*, 25(11): 2361-2369) [DOI: 10.11834/jig.200315]
- Kirkpatrick J, Pascanu R, Rabinowitz N, Veness J, Desjardins G, Rusu A A, Milan K, Quan J, Ramalho T, Grabska-Barwinska A, Hassabis D, Clopath C, Kumaran D and Hadsell R. 2017. Overcoming catastrophic forgetting in neural networks. *Proceedings of the National Academy of Sciences of the United States of America*, 114(13): 3521-3526 [DOI: 10.1073/pnas.1611835114]
- Ledoit O and Wolf M. 2004. A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis*, 88(2): 365-411 [DOI: 10.1016/S0047-259X(03)00096-4]
- Lee K, Lee K, Lee H and Shin J. 2018. A simple unified framework for detecting out-of-distribution samples and adversarial attacks//*Proceedings of the 32nd Conference on Neural Information Processing Systems*. Montreal, Canada: Curran Associates Inc.: 7167-7177 [DOI: 10.5555/3327757.3327819]
- Li C L, Sohn K, Yoon J and Pfister T. 2021. CutPaste: self-supervised learning for anomaly detection and localization//*Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Nashville, USA: IEEE: 9664-9674 [DOI: 10.1109/CVPR46437.2021.00954]
- Li W B, Xiong Y K, Fan Z C, Deng B, Cao F Y and Gao Y. 2024. Advances and trends of continual learning. *Journal of Computer Research and Development*, 61(6): 1476-1496 (李文斌, 熊亚锐, 范祉辰, 邓波, 曹付元, 高阳. 持续学习的研究进展与趋势. *计算机研究与发展*, 61(6): 1476-1496) [DOI: 10.7544/issn1000-1239.202220820]
- Li W J, Zhan J W, Wang J B, Xia B Z, Gao B B, Liu J, Wang C J and Zheng F. 2022. Towards continual adaptation in industrial anomaly detection//*Proceedings of the 30th ACM International Conference on Multimedia*. Lisboa, Portugal: ACM: 2871-2880 [DOI: 10.1145/3503161.3548232]
- Lian D Z, Zhou D Q, Feng J S and Wang X C. 2022. Scaling and shifting your features: a new baseline for efficient model tuning//*Proceedings of the 36th International Conference on Neural Information Processing Systems*. New Orleans, USA: Curran Associates Inc: #9 [DOI: 10.5555/3600270.3600279]
- Mehta S V, Patil D, Chandar S and Strubell E. 2023. An empirical investigation of the role of pre-training in lifelong learning. *The Journal of Machine Learning Research*, 24(1): #214 [DOI: 10.5555/3648699.3648913]
- Ni Y M and Chen S C. 2022. Continual unsupervised anomaly detection. *Scientia Sinica Informationis*, 52(1): 75-85 (倪一鸣, 陈松灿. 2022. 连续无监督异常检测. *中国科学: 信息科学*), 52(1): 75-85 [DOI: 10.1360/SSI-2021-0192]
- Rebuffi S A, Kolesnikov A, Sperl G and Lampert C H. 2017. iCaRL: incremental classifier and representation learning//*Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu, USA: IEEE: 5533-5542 [DOI: 10.1109/CVPR.2017.587]
- Reiss T, Cohen N, Bergman L and Hoshen Y. 2021. PANDA: adapting pretrained features for anomaly detection and segmentation//*Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Nashville, USA: IEEE: 2805-2813 [DOI: 10.1109/CVPR46437.2021.00283]
- Riemer M, Cases I, Ajemian R, Liu M, Rish I, Tu Y H and Tesauero G. 2018. Learning to learn without forgetting by maximizing transfer and minimizing interference [EB/OL]. [2025-01-15].
<https://arxiv.org/pdf/1810.11910.pdf>
- Rippel O, Mertens P and Merhof D. 2021. Modeling the distribution of normal data in pre-trained deep features for anomaly detection//*Proceedings of the 25th International Conference on Pattern Recognition*. Milan, Italy: IEEE: 6726-6733 [DOI: 10.1109/ICPR48806.2021.9412109]
- Tan Y W, Zhou Q H, Xiang X, Wang K, Wu Y C and Li Y B. 2024. Semantically-shifted incremental adapter-tuning is a continual ViT-ransformer//*Proceedings of 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Seattle, USA: IEEE: 23252-23262 [DOI: 10.1109/CVPR52733.2024.02194]
- Wang S Q, Cheng C, Shi M and Zhu D M. 2024. Defect detection method for industrial product surfaces with similar features by combining frequency and ViT. *Journal of Image and Graphics*, 29(10): 3074-3089 (王素琴, 程成, 石敏, 朱登明. 2024a. 结合频率和ViT的工业产品表面相似特征缺陷检测方法. *中国图象图形学报*, 29(10): 3074-3089) [DOI: 10.11834/jig.230532]

- Wang W P, Qin Y C and Shi W X. 2024. Review of unsupervised deep learning methods for industrial defect detection [J/OL]. [2024-11-01]. Journal of Computer Applications, 1-16 (王文鹏, 秦寅畅, 师文轩. 2024. 工业缺陷检测无监督深度学习方法综述 [J/OL]. [2024-11-01]. 计算机应用, 1-16
<http://kns.cnki.net/kcms/detail/51.1307.TP.20240911.1259.002.html>
- Yildiz O, Chan H, Raghavan K, Judge W, Cherukara M J, Balaprakash P, Sankaranarayanan S and Peterka T. 2022. Automated continual learning of defect identification in coherent diffraction imaging//2022 IEEE/ACM International Workshop on Artificial Intelligence and Machine Learning for Scientific Applications. Dallas, USA: IEEE: 1-6 [DOI: 10.1109/AI4S56813.2022.00006]
- Yin D S, Hu L Y, Li B and Zhang Y Q. 2023. Adapter is all you need for tuning visual tasks [EB/OL]. [2025-01-15].
<https://arxiv.org/pdf/2311.15010.pdf>
- Zhang G W, Wang L Y, Kang G L, Chen L and Wei Y C. 2023. SLCA: slow learner with classifier alignment for continual learning on a pre-trained model//Proceedings of 2023 IEEE/CVF International Conference on Computer Vision. Paris, France: IEEE: 19091-19101 [DOI: 10.1109/ICCV51070.2023.01754]
- Zhou D W, Cai Z W, Ye H J, Zhan D C and Liu Z W. 2024. Revisiting class-incremental learning with pre-trained models: Generalizability and adaptivity are all you need. International Journal of Computer, 133(3): 1012-1032 [DOI: 10.1007/s11263-024-02218-0]
- Zhou D W, Wang F Y, Ye H J and Zhan D C. 2023. Deep learning for class-incremental learning: a survey. Chinese Journal of Computers, 46(8): 1577-1605 (周大蔚, 汪福运, 叶翰嘉, 詹德川. 2023. 基于深度学习的类别增量学习算法综述. 计算机学报, 46(8): 1577-1605) [DOI: 10.11897/SP.J.1016.2023.01577]
- Zhou D W, Wang Q W, Ye H J and Zhan D C. 2023. A model or 603 exemplars: towards memory-efficient class-incremental learning [EB/OL]. [2025-01-15]. <https://arxiv.org/pdf/2205.13218.pdf>

作者简介

封筠,女,教授,主要研究方向为计算机视觉、机器学习和复杂网络分析。E-mail:fengjun@stdu.edu.cn

孟旭静,女,硕士研究生,主要研究方向为计算机视觉、连续学习和异常检测。E-mail:1202210012@student.stdu.edu.cn

尚玉全,男,硕士研究生,主要研究方向为计算机视觉和工业缺陷检测。E-mail:1202310021@student.stdu.edu.cn

牛超凡,男,工程师,主要研究方向为计算机视觉和工业缺陷检测。E-mail:niucf@wewinpe.com