CL-Anomaly: Layer-Adaptive Mixture-of-Experts with Multimodal Large Language Model for Continual Learning in Anomaly Detection

Wen Dong[1], Zhao Wang[2†], Shuangqing Zhang[1], Kai Sun[2], Ben Li[2], Guo-Sen Xie[3], Caifeng Shan[1], and Fang Zhao[1†]

[1] Nanjing University, Nanjing, China
[2] China Mobile Zijin Innovation Institute, Nanjing, China
[3] Nanjing University of Science and Technology, Nanjing, China
{wdong, shuangqing}@smail.nju.edu.cn
{wangzh8, sunk, liben}@js.chinamobile.com
{gsxiehm, caifeng.shan, zhaofang0627}@gmail.com

Abstract. Multimodal Large Language Models (MLLMs) excel in diverse vision tasks, but full-parameter retraining is computationally expensive as real-world knowledge evolves. Existing continual learning methods often suffer from semantic entanglement in parameter spaces across tasks, impeding the continuous deployment of models. This challenge is especially pronounced in Anomaly Detection (AD), which exhibits triple heterogeneity across modalities, domains, and defect scale variability, significantly complicating multi-task knowledge transfer. In this paper, we propose CL-Anomaly, a parameter-efficient fine-tuning framework based on an isolation-sharing collaboration to enable continual learning for anomaly detection with MLLMs. We introduce the task-private expert PrivLoRA, which physically isolates task-specific subspaces in the parameter space to prevent semantic entanglement of anomaly knowledge in diverse scenarios. The Layer-Adaptive Shared Experts maintain cross-task representations within a unified feature space, enabling knowledge sharing between previous and new tasks. Furthermore, we propose a Layer-Adaptive Knowledge Transfer strategy that automatically selects and dynamically updates the layer-wise key shared experts of each task via a momentum-based mechanism, promoting effective knowledge transfer across related anomaly detection tasks. Extensive experiments across three continual learning scenarios for anomaly detection, including class-incremental, cross-domain, and cross-modal, demonstrate that CL-Anomaly outperforms state-of-the-art methods. Code is available at `https://github.com/WenDongyp/CL-Anomaly`.

Keywords: Anomaly Detection · Continual Learning · Multimodal Large Language Model

† Corresponding authors.

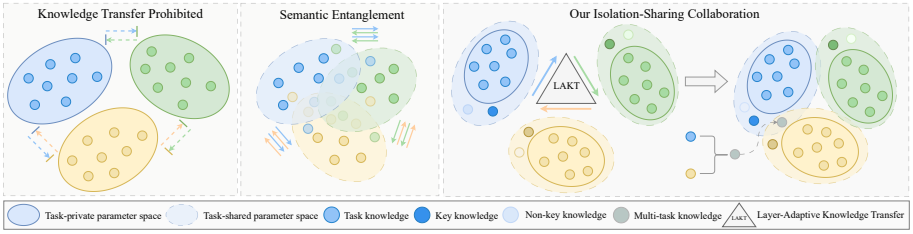
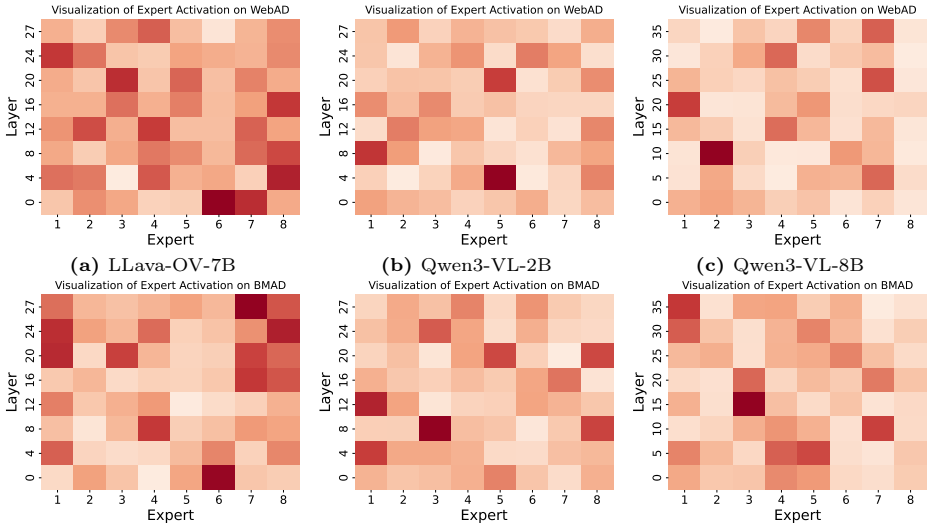


Fig. 1: (Left) Parameter isolation methods prohibit cross-task knowledge transfer, preventing knowledge complementarity. (Middle) Parameter sharing methods often lead to semantic entanglement, where parameter spaces of multiple tasks overlap and conflict. (Right) Our isolation–sharing collaboration framework maintains task-specific knowledge within isolated parameter spaces while enabling cross-task knowledge transfer through shared spaces. Furthermore, we introduce a layer-adaptive knowledge transfer strategy that preserves layer-wise key knowledge during parameter updates and facilitates the formation of multi-task knowledge, while non-key knowledge gradually fades throughout continual learning.

1 Introduction

Anomaly detection (AD) aims to identify samples that deviate from normal data distribution and has been widely applied in areas such as industrial quality control [7, 21, 48, 51] and medical imaging diagnosis [25, 57, 60, 61]. Conventional methods typically provide only anomaly scores, offer no user interaction, and are often restricted to a single modality (e.g., 2D or 3D), as well as specific task domains. In contrast, multimodal large language models (MLLMs) [1, 3, 11, 15, 29] inherently support cross-modal semantic understanding and human–computer interaction, allowing instruction prompts to guide the model in generating desired responses while handling inputs across multiple modalities and domains. Consequently, leveraging MLLMs for anomaly detection represents a promising direction. However, real-world anomaly detection is complex and evolving, making per-pipeline model training costly and impractical. Moreover, historical data privacy, commonly encountered in practical AD systems, restricts the application of many existing anomaly detection paradigms. Equipping MLLMs with continual learning capabilities enables incremental adaptation to new categories and domains, offering an effective solution. Hence, advancing continual learning for anomaly detection with MLLMs is of critical significance.

Parameter-efficient fine-tuning [22, 23, 28, 31, 62] is the most prevalent continual learning paradigm for MLLMs, introducing few trainable parameters for rapid task adaptation while keeping pretrained weights frozen. Nevertheless, it inevitably faces catastrophic forgetting, the classical challenge in continual learning, which is typically caused by semantic entanglement across tasks or the absence of knowledge transfer mechanisms. Recently, many efforts have been made to mitigate catastrophic forgetting. MoELoRA [9], SMoLoRA [56], and CL-MoE [24] update parameters within a unified representation space, enabling positive knowledge transfer across similar tasks. O-LoRA [55], MoA [14], and



(a) LLava-OV-7B **(b)** Qwen3-VL-2B **(c)** Qwen3-VL-8B

(d) LLava-OV-7B **(e)** Qwen3-VL-2B **(f)** Qwen3-VL-8B

Fig. 2: Heatmaps of expert activation across different layers in the LoRA-based MoE architecture on WebAD and BMAD datasets. Layer-wise experts exhibit varying activation preferences across tasks, which is widely observed across MLLMs of diverse scales and architectures, highlighting the importance of differentiating layer-wise experts based on their task preferences.

HiDe-LLaVA [20] isolate parameter spaces for individual tasks, effectively preventing semantic entanglement. However, relying solely on either approach entails inherent limitations: shared parameters risk conflict when tasks span diverse domains, while isolated parameters impede knowledge transfer across tasks, as illustrated in Fig. 1 (Left, Middle).

In this paper, we propose CL-Anomaly to bridge the research gap in continual learning for anomaly detection with MLLMs. As shown in Fig. 1 (Right), CL-Anomaly incorporates task-private and task-shared paths, with Layer-Adaptive Knowledge Transfer executed on the shared parameter space. In the task-private path, each task employs a dedicated LoRA expert, termed PrivLoRA, which updates only within its own task. Fully isolating the parameter spaces of PrivLoRAs inherently mitigates semantic interference in anomaly knowledge across modalities and domains. The task-shared path enables cross-task knowledge transfer by jointly optimizing multiple LoRAs in a unified parameter space. Together, the two paths effectively mitigate catastrophic forgetting. Moreover, we further observe that in the LoRA-based MoE architecture, the activity of individual LoRA experts e at each layer l varies across different tasks t, as illustrated in Fig. 2. Based on this observation, we propose Layer-Adaptive Knowledge Transfer, a simple yet efficient strategy to preserve the knowledge of layer-wise key shared experts. During training, it automatically identifies highly active experts for the current task at zero extra cost and prioritizes their information when up-

dating shared parameters via dynamic momentum merging, thereby preserving core anomaly knowledge while facilitating knowledge integration across related anomaly detection tasks. In summary, our contributions are as follows:

- We introduce CL-Anomaly, which adopts a *private-shared* dual-path fine-tuning framework to prevent semantic entanglement while enabling knowledge transfer. To our knowledge, we are the first to bring the continual learning paradigm of MLLMs into anomaly detection.
- We found significant differences in task preferences among experts at different layers and propose a Layer-Adaptive Knowledge Transfer strategy to adaptively retain layer-wise key expert knowledge while facilitating cross-task knowledge transfer.
- Extensive experiments on three continual learning scenarios for anomaly detection, including class-incremental, cross-domain, and cross-modal, show that CL-Anomaly effectively mitigates catastrophic forgetting and outperforms current state-of-the-art methods.

2 Related Work

Anomaly Detection. Early anomaly detection methods relied mainly on reconstruction error [19,36,38] or feature matching [42,48,53] to detect anomalies, following a *one-class-one-model* training paradigm, which is highly impractical in real-world scenarios. In recent years, researchers have focused on few-shot [26, 32, 50] and zero-shot [8, 13, 64] learning paradigms, leveraging visual-language models with strong generalization capabilities (e.g., CLIP) to calculate anomaly scores by comparing visual and textual feature representations, but their detection capability is often constrained by limited exposure to the target dataset. Real-world scenarios, including the continual emergence of new product categories in industry and cross-domain transfer from industrial inspection to medical imaging, exemplify the significance of continual anomaly detection, yet few studies have investigated it. Accordingly, we propose the CL-Anomaly framework, enabling efficient continual learning for anomaly detection across both class and domain incremental scenarios.

Multimodal Large Language Models. With the rapid development of large language models (LLMs) [2, 52], growing attention has been devoted to equipping them with multimodal learning capabilities, thereby extending their applications to complex scenarios. Some studies integrate image encoders with modality connectors into LLMs to enable visual understanding, gradually expanding to additional modalities such as video and audio. These multimodal large language models (MLLMs) demonstrate outstanding capabilities in instruction following [17, 39], knowledge reasoning [10, 16, 63] and various vision downstream tasks [35, 54, 58]. However, research on MLLMs for anomaly detection remains scarce. This paper aims to leverage MLLMs for continual learning in anomaly detection, providing a novel insight and method for multimodal anomaly detection.

Continual Learning. In real-world, the continuous evolution of product categories and task types is widespread. Accordingly, MLLMs are expected to dynamically acquire new knowledge while robustly retaining previously learned information. Recent continual learning methods can be mainly categorized into replay-based, regularization-based, and parameter-efficient fine-tuning methods. Replay-based methods [43, 47, 65], preserve a shared pre-trained representation space by storing and replaying knowledge from previous tasks. However, as the accumulated knowledge grows, these methods incur substantial storage overhead and are constrained in anomaly detection scenarios involving data privacy. Regularization-based methods [12, 27, 34] preserve the stability of previous task parameters during incremental learning through explicit constraints on critical model parameters or the loss function. Parameter-efficient fine-tuning methods achieve rapid adaptation to new tasks by updating a limited number of additional parameters while keeping the pre-trained backbone frozen to preserve the model's original knowledge. MoELoRA [9] learns diverse knowledge representations through the collaboration of multiple LoRA experts. CL-MoE [24] employs a dynamic expert update strategy, achieving a trade-off between plasticity and stability. But the absence of parameter decoupling among experts renders CL-MoE susceptible to semantic entanglement. Our method combines parameter-decoupled private experts with shared experts, maintaining task-specific knowledge while promoting cross-task knowledge transfer. HiDe-LLaVA [20] introduces task-specific expansion and task-general fusion but treats all non-top-layer LoRA experts uniformly, neglecting the layer-wise differences in expert task sensitivity. Our method explicitly accounts for layer-wise expert task preferences, dynamically identifying layer-wise key experts salient to the current task and effectively preserving their knowledge during transfer.

3 Preliminary

Continual Learning Setup. Our work focuses on continual learning for anomaly detection with MLLM, aiming to sequentially learn tasks while mitigating catastrophic forgetting. Given T tasks, continual instruction tuning sequentially accesses the data of each task $\mathcal{D}_t = \{V_i^t, Q_i^t, A_i^t\}_{i=1}^{N^t},\ t \in \{1, 2, \ldots, T\}$, where N^t denotes the data size of task t. The MLLM is tasked with learning the current task t while preserving the knowledge from previous tasks $t-1$.

Low-Rank Adaptation. Low-Rank Adaptation (LoRA) is a parameter-efficient fine-tuning method. By freezing the pretrained weights $W_0 \in \mathbb{R}^{d\times k}$ and introducing a trainable low-rank update $\Delta W = BA$, where $B \in \mathbb{R}^{d\times r}$, $A \in \mathbb{R}^{r\times k}$, and the rank $r \ll \min(d, k)$, it adapts the model to new tasks while significantly reducing the number of updated parameters. It is typically injected into the linear projection layers of the self-attention and feed-forward networks. The forward pass with LoRA can be expressed as:

$$f(x) = W_0 x + \Delta W x = W_0 x + \frac{\alpha}{r} B\,(Ax) \tag{1}$$

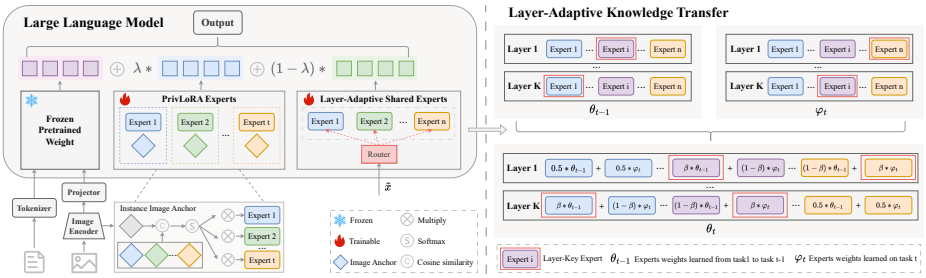


Fig. 3: Overview of our method. (Left) CL-Anomaly consists of parallel task-private and task-shared paths. PrivLoRA is updated only within its task, extracting visual anchors with a multimodal encoder and assigning expert weights by similarity to task anchors. The task-shared path, termed Layer-Adaptive Shared Experts, dynamically selects experts with a gating function and updates parameters using a layer-adaptive strategy. (Right) Layer-Adaptive Knowledge Transfer: identifies key layer-wise experts from gating probability vectors at zero extra cost and prioritizes key expert representations during dynamic momentum merging.

where α denotes the scaling factor controlling the magnitude of the low-rank update.

4 Method

4.1 Framework Overview

In this work, we propose CL-Anomaly, a dual-path mixture-of-experts framework for continual learning in anomaly detection. The overall architecture is illustrated in Fig. 3. CL-Anomaly introduces parallel task-private and task-shared expert paths, enabling simultaneous semantic decoupling and knowledge transfer. During training, only current PrivLoRA and task-shared experts are updated, while other private experts remain frozen. At the inference stage, the model combines adapted increments from both expert paths with frozen weights output:

$$h = W_0 x + \lambda f_p(x) + (1 - \lambda) f_s(x) \tag{2}$$

where λ denotes the weighting coefficient, fixed at 0.5. $f_p(\cdot)$ and $f_s(\cdot)$ denote the outputs of the private and shared paths, with further discussion in Sections 4.2 and 4.3.

To further mitigate catastrophic forgetting, we propose a layer-adaptive knowledge transfer strategy that identifies key experts layer by layer and dynamically preserves their knowledge during parameter updates, enabling efficient layer-wise task-critical knowledge transfer.

4.2 PrivLoRA Experts

Conventional LoRA-based fine-tuning approaches are inherently susceptible to semantic entanglement in multi-task continual learning. To overcome this limita-

tion, inspired by HiDe-LLaVA [20], we further introduce task-private PrivLoRA experts. Each task is assigned a dedicated PrivLoRA expert, receiving gradient updates exclusively during the training of the corresponding task. PrivLoRA experts maintain parameter spaces fully isolated from other tasks, preventing cross-task interference in the shared representation space and eliminating semantic entanglement at its source. We use the visual encoder of LLaVA to extract the visual anchor $anchor_t$ for task t during training:

$$anchor_t = \frac{1}{N_t}\sum_{i=1}^{N_t}(\frac{1}{M}\sum_{m=1}^{M}[f_v(x_i^t)]_m) \tag{3}$$

where x_i^t denotes the input image instance i of task t, f_v is the visual encoder of LLaVA, and $f_v(x) \in \mathbb{R}^{B\times M\times C}$ denotes the patch features of input x, with B the batch size, M the number of patch tokens, and C the feature dimension.

The PrivLoRA expert appropriate for the input image x_I is selected based on the cosine similarity between its instance visual anchor $anchor_I$ and the stored task anchors. Specifically, we quantify the correlation between the input instance visual anchor $anchor_I$ and the stored anchor of task t as μ_t:

$$\mu_t = \frac{\exp\big(\mathrm{Cos}(anchor_I,\, anchor_t)\big)}{\sum_{j=1}^{T}\exp\big(\mathrm{Cos}(anchor_I,\, anchor_j)\big)}, \quad t = 1,\ldots,T \tag{4}$$

where $\mathrm{Cos}(.,.)$ represents the cosine similarity. The output of the PrivLoRA experts path is given as follows:

$$f_p(\tilde{x_I}) = \sum_{i=1}^{t}\mu_i P_i(\tilde{x}_I) \tag{5}$$

where $\tilde{x_I}$ denotes the intermediate representation of the input instance x_I within the large language model block, and $P_i(\cdot)$ represents the i-th PrivLoRA expert.

4.3 Layer-Adaptive Shared Experts

Task-Shared Experts. Although physically isolated task-private experts effectively eliminate semantic entanglement, the complete decoupling of parameters simultaneously hinders knowledge transfer across tasks. Prior studies [37, 45, 46] have demonstrated that knowledge transfer supports the progressive accumulation of information, with new task knowledge complementing and reinforcing the representations of previous tasks, especially when there are intrinsic semantic or distributional correlations, e.g., in class-incremental learning under a shared imaging environment. We thus employ a task-shared expert path to complement PrivLoRA, addressing its constraints in cross-task knowledge transfer. The task-shared path adopts a vanilla mixture-of-experts LoRA architecture, treating each LoRA as an individual expert network. In this architecture, the router mapping the input x to expert assignment weights ω:

$$\omega_e(x) = \frac{\exp(g_e \cdot x)}{\sum_{j=1}^{n}\exp(g_j \cdot x)}, \quad e = 1,\ldots,n \tag{6}$$

where g_e denotes the trainable gating parameter of expert e and n is the number of shared experts. The output of the task-shared path is obtained by weighting the experts accordingly:

$$f_s(\tilde{x}) = \sum_{e=1}^{n} \omega_e(\tilde{x}) \cdot F_e(\tilde{x}) \tag{7}$$

where $F_e(\cdot)$ denotes the shared LoRA expert e.

Layer-Adaptive Knowledge Transfer. We further observe significant differences in task sensitivity among experts e at each layer l for different tasks t within the LoRA-based MoE architecture, with individual experts showing varying activation in fine-grained tasks and global understanding. Therefore, we introduce a Layer-Adaptive Knowledge Transfer strategy to identify key layer-wise LoRA experts and enhance cross-task knowledge propagation. Specifically, we compute the task relevance score S of shared experts per layer during training by accumulating gating probability vectors.

$$S_t^{(l,e)} = \frac{1}{|\mathcal{D}_t|} \sum_{x \in \mathcal{D}_t} \omega_{l,e}(x) \cdot \mathbb{I}\left[e = \underset{j \in \{1,\ldots,n\}}{\operatorname{argmax}} \ \omega_{l,j}(x) \right] \tag{8}$$

where $\omega_{l,e}(x)$ denotes the routing probability of input sample x to expert e at layer l. The top-K shared experts with the highest scores $S_t^{(l,e)}$ are designated as key LoRA experts $e_{t,l}^*$, remaining consistently active at layer l for task t and demonstrating strong task relevance. This is a simple yet efficient strategy that selects key experts with minimal computational overhead, without requiring additional forward or backward propagation. Let $E_t^* = \{ e_{t,l}^* \mid l = 1, 2, \ldots, L \}$ denote the set of layer-wise key experts for task t, where L is the number of layers in the LLM. We partition all shared experts into three mutually exclusive subsets: $E_t^{only} = \{e \in E_t^* \mid e \notin E_{t-1}^*\}$, the experts key only to task t; $E_{t-1}^{only} = \{e \in E_{t-1}^* \mid e \notin E_t^*\}$, the experts key only to task $t-1$; and the remaining experts. The transfer coefficient of shared expert e at layer l is defined as:

$$\beta_{l,e} = \begin{cases} \beta, & \text{if } e \in E_t^{only} \\ 1-\beta, & \text{if } e \in E_{t-1}^{only} \\ 0.5, & \text{otherwise} \end{cases} \tag{9}$$

Let θ_{t-1} denote the shared expert parameters learned by the model from tasks 1 to $t-1$, and let φ_t represent the parameters initialized with θ_{t-1} and fine-tuned on task t. We perform layer-adaptive momentum merging of θ_{t-1} and φ_t to enable efficient knowledge transfer across tasks, thereby maximizing the update of key knowledge for the new task while retaining knowledge from previous tasks. The process can be expressed as:

$$\theta_t^{(l,e)} = (1 - \beta_{l,e}) \cdot \theta_{t-1}^{(l,e)} + \beta_{l,e} \cdot \varphi_t^{(l,e)} \tag{10}$$

where θ_t denotes the updated shared expert parameters for task t. The layer-adaptive knowledge transfer strategy retains key expert knowledge across all

layers for each task, facilitating knowledge transfer while effectively mitigating catastrophic forgetting.

5 Experiments

5.1 Experimental Setup

Datasets. We conduct experiments on Anomaly-Instruct-125K [59], the largest anomaly-detection-related visual instruction question-answering dataset to date. Spanning 2D to 3D data and covering industrial and medical domains, Anomaly-Instruct-125K is composed of classical datasets including MVTec-AD [5], MVTec-3D [6], Real3D-AD [41], Anomaly-ShapeNet [30], and BMAD [4], along with WebAD [59], an in-the-wild multi-class image dataset. We create three subsets from WebAD based on object categories, each containing 35 classes, to evaluate model performance in class-incremental learning scenarios. BMAD, a medical-domain dataset, is used to assess model performance under cross-domain continual learning. Additionally, 3D datasets are employed to evaluate model capability in continual learning across modalities. See Appendix B for more dataset details.
Evaluation Metrics. Following prior continual learning works [9, 18, 20, 24], we evaluate continual learning performance using *Avg*, *Last*, and BWT metrics based on per-task accuracy across stages. Given the inherent class imbalance in anomaly detection datasets, we further report *Accuracy*, *Precision*, *Recall*, and $F1$-*score* to ensure a fair and comprehensive evaluation of continual anomaly detection performance. Detailed definitions and metric selection criteria are provided in Appendix C.1.
Baselines. We compare CL-Anomaly with existing continual learning baseline methods. LoRA [23] is a standard parameter-efficient fine-tuning approach that is widely adopted for MLLM adaptation. MoELoRA [9] leverages a mixture of LoRA experts to learn diverse knowledge representations. CL-MoE [24] employs dynamic updates of a mixture of experts to balance expert workload. HiDe-LLaVA [20] combines non-top-layer experts while preserving the independence of top-layer experts. We also report performance under zero-shot and multi-task fine-tuning settings, where all tasks are learned simultaneously in the multi-task setup without task-ordering constraints. The detailed of each method can be found in Appendix C.2.
Implementation details. We use LLaVA-OV-7B [29] as the base MLLM, keeping all its parameters frozen, and embed LoRA experts into the linear layers of the language model following the standard LoRA fine-tuning protocol employed by LLaVA. We set the number of shared experts n to 4, with $k = 1$ key expert per layer, the LoRA rank $r = 36$, the scaling factor $\alpha = 72$, and the momentum merging weight $\beta = 0.75$. All tasks are trained for 1 epoch using the AdamW [44] optimizer, with a learning rate of 1×10^{-5} and a warm-up ratio of 0.03. For fair comparison, all baseline methods use the same base model and fine-tuning strategy with carefully tuned hyperparameters. All experiments are conducted on 8 RTX A6000 GPUs. Further details are provided in Appendix C.3.

Table 1: Comparison of our method with various approaches on continual learning for anomaly detection. The best results are highlighted in **bold**, and the second best results are underlined.

	Method	WebAD-1	WebAD-2	WebAD-3	Mvtec-AD	BMAD	Mvtec-3D	Real3D-AD	Anomaly-Shapenet	*Average*
	Zero-shot	86.17	86.44	84.54	92.05	61.53	69.06	55.58	56.62	73.99
	Multi-task	93.41	92.47	93.48	97.73	92.59	83.92	77.69	75.71	88.38
Avg	LoRA	88.59	87.63	86.13	94.14	73.97	80.87	72.91	**76.36**	82.58
	MoELoRA	89.79	88.85	88.22	94.94	74.59	77.47	62.65	71.82	81.04
	CL-MoE	90.74	89.92	89.43	94.96	76.64	76.67	59.76	66.88	80.63
	HiDe-LLaVA	83.67	84.06	82.48	91.44	89.18	74.96	56.47	61.04	77.91
	CL-Anomaly	**92.59**	**91.59**	**92.20**	**95.63**	**89.44**	**81.30**	**74.30**	73.64	**86.34**
Last	LoRA	81.58	82.36	80.28	92.43	60.94	79.00	70.92	**76.36**	77.98
	MoELoRA	86.98	85.96	85.93	93.40	65.41	74.68	58.57	71.82	77.84
	CL-MoE	88.60	87.95	87.49	93.73	69.29	73.73	56.97	66.88	78.08
	HiDe-LLaVA	75.31	76.39	73.47	89.13	**87.88**	68.54	54.78	61.04	73.32
	CL-Anomaly	**92.06**	**90.37**	**91.75**	**94.05**	**87.88**	**81.24**	**73.90**	73.64	**85.61**
BWT	LoRA	-5.13	-5.59	-8.76	-4.42	-24.36	-3.02	-3.98	-	-7.89
	MoELoRA	-2.72	-2.22	-4.23	-3.08	-22.43	-4.62	-8.16	-	-6.78
	CL-MoE	-1.87	-1.90	-3.48	-2.99	-19.53	-5.96	-5.58	-	-5.90
	HiDe-LLaVA	-6.62	-7.36	-10.65	-5.22	-4.55	-6.70	-3.39	-	-6.36
	CL-Anomaly	**-0.07**	**-0.53**	**-0.79**	**-1.81**	**-2.95**	**0.21**	**-0.80**	-	**-0.96**

5.2 Comparison with the State-of-the-art

Table 1 presents a comparison of our method with the baselines. Based on these results, we provide the following key findings:

- CL-Anomaly exhibits outstanding overall performance, outperforming the current state-of-the-art. Compared with CL-MoE, the second-best method, it improves Avg, Last, and BWT by 5.68%, 7.44%, and 4.94%, respectively. Notably, CL-Anomaly approaches the upper bound of Avg and Last metrics in multi-task settings, demonstrating both its effectiveness and robust resistance to forgetting.
- Parameter sharing and isolation strategies exhibit advantages in different scenarios. Sharing-based methods (e.g., MoELoRA, CL-MoE) achieve strong performance on class-incremental tasks (WebAD-1 to WebAD-3) by effectively transferring knowledge among similar tasks. However, their transferability is limited in cross-domain increments, such as the medical BMAD dataset, due to substantial domain differences. Conversely, isolation-based methods like HiDe-LLaVA achieve modest performance on class-incremental tasks but remain stable in cross-domain scenarios, highlighting the rationale of CL-Anomaly's isolation–sharing collaborative framework. Additional analyses of the baselines are provided in Appendix C.3.
- Most baseline methods perform poorly on 3D datasets due to large representation gaps between 2D and 3D tasks, causing severe semantic shifts. HiDe-LLaVA is constrained by its inference mechanism, which integrates non-top-level parameters across domains and modalities, causing knowledge entanglement and limiting performance. In contrast, CL-Anomaly leverages fully isolated private expert paths and a layer-adaptive knowledge transfer strategy to preserve the knowledge of each modality task and mitigate the impact of modality differences.

Table 2: Comprehensive metrics comparison in the last stage.

Method	Accuracy	Precision	Recall	F1-score	Average
Zero-shot	73.99	80.46	65.39	69.52	72.34
Multi-task	88.38	90.13	83.83	86.18	87.13
LoRA	77.98	87.65	66.16	74.74	76.63
MoELoRA	77.84	87.41	64.84	72.95	75.76
CL-MoE	78.08	85.51	66.30	73.47	75.84
HiDe-LLaVA	73.32	71.14	67.75	68.61	70.21
CL-Anomaly	**85.61**	**89.61**	**78.00**	**82.51**	**83.93**

Table 3: Ablation study of proposed CL-Anomaly.

Method	Avg	Last	BWT
Full Model	86.34	85.61	-0.96
w/o. PrivLoRA	83.58	81.60	-3.70
w/o. SharedLoRA	84.97	82.63	-3.29
w/o. Layer-Adaptive	85.09	84.06	-1.38

Table 4: Quantitative comparison of anomaly reasoning capabilities.

Method	Mvtec-AD			BMAD			Real3D-AD		
	ROUGE-L	SBERT	LLM-Score	ROUGE-L	SBERT	LLM-Score	ROUGE-L	SBERT	LLM-Score
Zero-shot	41.00	79.81	6.50	31.73	68.67	5.02	34.38	73.53	4.67
CL-MoE	43.47	83.81	7.17	38.24	73.25	6.29	41.68	82.03	5.94
CL-Anomaly	**44.65**	**85.40**	**7.64**	**40.70**	**76.94**	**6.77**	**43.14**	**83.27**	**6.60**

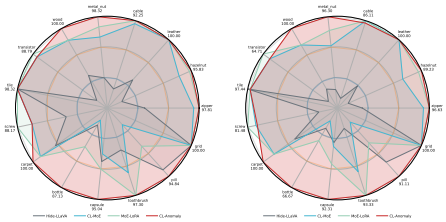
Given the inherent class imbalance in anomaly detection datasets and to ensure a fair comparison with existing continual learning approaches, we further report the *Accuracy*, *Precision*, *Recall*, and *F1-score* obtained after completing all training stages. As shown in Table 2, CL-Anomaly achieves superior performance in all metrics, demonstrating the strong practicality of our method in anomaly detection tasks. More details and complete metrics for each task can be found in Appendix D.1.

5.3 Ablation Study

We conducted a series of ablation experiments to verify the effectiveness of each component in CL-Anomaly, where Layer-Adaptive denotes the Layer-wise Adaptive Knowledge Transfer. As shown in the Table 3, each component contributes to the overall performance gain. Removing PrivLoRA leads to a significant performance drop, highlighting the importance of task-specific parameter independence for continual learning in anomaly detection. Disabling the shared path (SharedLoRA) restricts knowledge flow among similar tasks and results in notable degradation when the visual-anchor-based routing of PrivLoRA is suboptimal. The layer-adaptive knowledge transfer mechanism further promotes knowledge sharing across tasks while preserving layer-wise key information, effectively mitigating catastrophic forgetting.

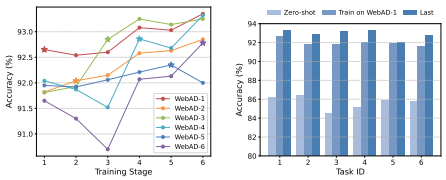
5.4 Further Analysis

Evaluation on Anomaly Reasoning. To further validate the performance of our model in anomaly reasoning under continual learning, we evaluated CL-Anomaly and CL-MoE on three representative datasets: the industrial domain



(a) Accuracy (b) F1-score

Fig. 4: Visualization of CL-Anomaly, and the SOTA continual learning approaches under class-incremental learning on practical single-anomaly scenarios.



(a) (b)

Fig. 5: (a) Accuracy of each task across stages in CIL, ⋆ indicates the target task being learned at that stage. (b) Accuracy under CIL compared to zero-shot.

(Mvtec-AD), the medical domain (BMAD), and the 3D modality (Real3D-AD), reporting results after all tasks were learned. We evaluate reasoning performance using Rouge-L, SBERT, and LLM-Score metrics. Specifically, Rouge-L evaluates the structural and lexical consistency between model outputs and reference answers; SBERT evaluates the semantic similarity under varied linguistic formulations; and LLM-Score leverages a large language model as an automatic judge to compare candidate and reference responses, providing a comprehensive measure of correctness, completeness, and reasoning quality. As shown in Table 4, our method outperforms the current continual learning SOTA, CL-MoE, in terms of expression fidelity, semantic consistency, and holistic reasoning quality, demonstrating that CL-Anomaly preserves robust anomaly reasoning capabilities for previously learned tasks even after multi-stage training. More details are provided in Appendix D.2.

Experiments in Traditional Continuous AD Scenarios. Tables 1 and 2 demonstrate the continual anomaly detection capability of our method across multiple complex scenarios. We further validate CL-Anomaly's continual learning performance under single-anomaly scenarios, which are prevalent in practical continual AD systems. Following the experimental setting of traditional continual anomaly detection works [40, 49], we partition the Mvtec-AD dataset into 15 category-based subsets and learn each subtask sequentially. We evaluate the model performance after training on all subtasks. As shown in Fig. 4, CL-Anomaly maintains superior detection capability despite multi-stage training on non-target category tasks, demonstrating the effectiveness of our method in practical continual anomaly detection systems. Detailed metrics and further descriptions are provided in Appendix D.3.

Analysis of Knowledge Transfer in CIL. We further introduce additional data categories from the WebAD dataset to construct a larger-scale class incremental learning (CIL) scenario, which is the most common task setting in real-world anomaly detection. Unlike the main experiments, this study reports task accuracy at each stage, regardless of whether the model has encountered the corresponding task. As shown in Fig. 5a, task performance generally improves

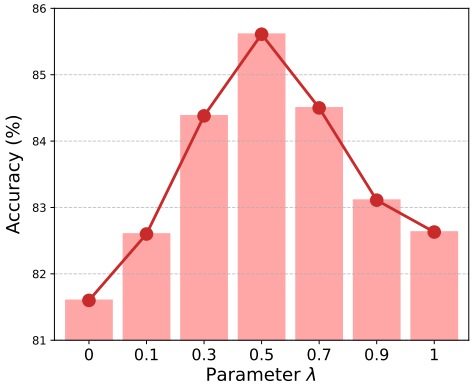


Fig. 6: Average last accuracy of λ.

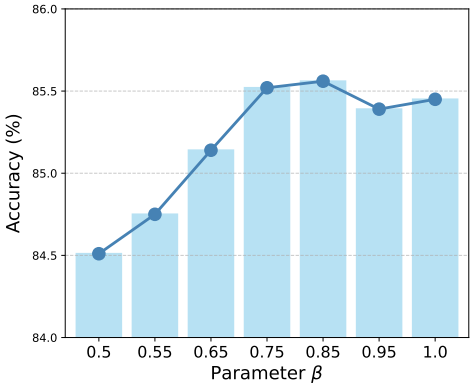


Fig. 7: Average last accuracy of β.

Table 5: Zero-shot results. *multi* denotes multi-task joint training.

Method	VisA	DTD
IAD-R1-7B (LLaVA)	76.09	91.18
IAD-R1-7B (Qwen)	65.47	77.91
AnomalyOV-7B	78.53	94.33
Ours-7B	75.41	95.86
Ours-multi-7B	76.43	96.63

Table 6: Analysis on the number of shared experts.

Method	$\langle n, k \rangle$	Avg	Last	BWT
CL-Anomaly	$\langle 1, 1 \rangle$	85.30	82.29	-3.68
	$\langle 2, 1 \rangle$	86.60	85.43	-1.68
	$\langle 4, 1 \rangle$	86.31	85.52	**-0.96**
	$\langle 8, 2 \rangle$	**86.75**	**85.87**	-1.15

Table 7: Performance under different task orders.

Method	Forward			Random		
	Avg	Last	BWT	Avg	Last	BWT
CL-MoE	80.63	78.08	-5.90	81.18	77.99	-5.70
CL-Anomaly	86.34	85.61	-0.96	84.81	84.84	-0.72

throughout continual learning, even after the task has been fully trained, indicating that CL-Anomaly enables positive backward knowledge transfer, where learning new tasks boosts the performance of previously learned ones. Moreover, we compare task accuracy under three conditions: zero-shot, after training only on WebAD-1, and after completing all tasks, as shown in Fig. 5b. After training on only the first task, CL-Anomaly markedly surpasses the zero-shot baseline across all tasks, demonstrating strong forward knowledge transfer, where learning early tasks improves performance on subsequently unseen tasks.

Performance on ZSAD. We further compare CL-Anomaly with powerful zero-shot anomaly detection MLLMs [33,59] under the zero-shot setting. Specifically, we evaluate on the DTD and VisA subsets (1,000 anomalous and 761 normal images). The results are presented in Table 5. Despite not being explicitly designed for zero-shot generalization, CL-Anomaly achieves competitive performance against these specialized ZSAD methods, highlighting its strong inherent transferability.

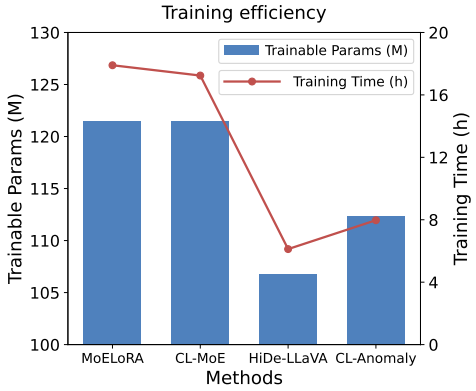


Fig. 8: Training efficiency comparison of our method and representative approaches.

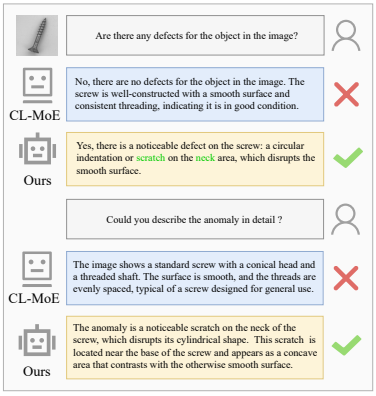


Fig. 9: Visualization of proposed CL-Anomaly and CL-MoE.

Impact of Hyperparameter λ. Fig. 6 illustrates the impact of the hyperparameter λ in Eq.2. Both excessively small and large values of λ degrade performance, while $\lambda = 0.5$ strikes an optimal balance, effectively mediating the trade-off between shared and private experts.

Impact of Hyperparameter β. Fig. 7 shows the impact of hyperparameter β in Eq.9. Larger β strengthens key knowledge retention and fades irrelevant knowledge; moderate increases improve performance, yet β near 0.5 degenerates into simple fusion and performance declines, demonstrating LAKT's robustness for β within a reasonable range. And considering the relative balance between prior and novel knowledge, we set the default $\beta_{l,e}$ to 0.5.

Impact of Shared Expert Number. Table 6 presents the impact of the number of shared experts n and key experts per layer k on model performance. CL-Anomaly maintains stable performance across a reasonable parameter range, and even with only two shared experts, it achieves competitive results, enabling efficient cross-task knowledge transfer with minimal additional shared experts.

Impact of Task Order. We further investigate the impact of different task orders on CL-Anomaly performance. Specifically, we randomly generated the following task order: WebAD-3 → Real3D-AD → WebAD-1 → Anomaly-Shapenet → MvTec-AD → WebAD-2 → MvTec-3D → BMAD, and compared the results with CL-MoE, as shown in the Table 7. Although the order of knowledge transfer across tasks inevitably affects performance, CL-Anomaly exhibits relatively stable performance and outperforms the state-of-the-art method CL-MoE.

Analysis of Training Parameters and Time. Fig. 8 presents the trainable parameters and training time of CL-Anomaly compared with baseline methods. MoELoRA and CL-MoE incur the highest training cost, while CL-Anomaly achieves comparable efficiency to HiDe-LLaVA with markedly improved performance. Compared with the pretrained backbone (8020.92 M), CL-Anomaly

incurs only a 1.4% parameter overhead, demonstrating superior parameter efficiency and deployment feasibility.

Example Analysis. Fig. 9 shows an example from industrial scenario. CL-Anomaly accurately detects anomalies while providing clear visual evidence and precise descriptions of anomaly regions, demonstrating strong anomaly understanding and reasoning capabilities. Meanwhile, the support for multi-round human–machine interaction makes it more user-friendly than traditional detection models. More examples are provided in Appendix F.

6 Conclusion

In this paper, we propose the CL-Anomaly framework, which first introduces continual learning with MLLM into anomaly detection task. We design a dual-path architecture combining parameter decoupling and shared experts to balance task-specific knowledge retention and cross-task knowledge transfer. The parameter spaces of the private expert PrivLoRA are isolated, with only the task-specific low-rank matrices updated. Shared experts enable multiplexing of multi-task knowledge through sparse gating within a unified representation space. We further observed that experts at different layers show distinct task preferences, motivating our Layer-Adaptive Knowledge Transfer strategy, which automatically identifies key experts at each layer with zero additional cost and enhances their contribution through dynamic momentum merging. This approach preserves critical information and promotes knowledge transfer, effectively mitigating catastrophic forgetting. Extensive experiments demonstrate that CL-Anomaly consistently outperforms SOTA methods across class-incremental, cross-domain, and cross-modal continual learning scenarios in anomaly detection tasks.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (No. 62476124, 62276134), the Fundamental and Interdisciplinary Disciplines Breakthrough Plan of the Ministry of Education of China (JYB2025XDXM902), the Natural Science Foundation of Jiangsu Province (No. BK20242015), the Gusu Innovation and Entrepreneur Leading Talents (No. ZXL2025322), and Nanjing University-China Mobile Communications Group Co.,Ltd. Joint Institute.

References

1. Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F.L., Almeida, D., Altenschmidt, J., Altman, S., Anadkat, S., et al.: Gpt-4 technical report. arXiv preprint arXiv:2303.08774 (2023)
2. Bai, J., Bai, S., Chu, Y., Cui, Z., Dang, K., Deng, X., Fan, Y., Ge, W., Han, Y., Huang, F., et al.: Qwen technical report. arXiv preprint arXiv:2309.16609 (2023)

3. Bai, S., Chen, K., Liu, X., Wang, J., Ge, W., Song, S., Dang, K., Wang, P., Wang, S., Tang, J., et al.: Qwen2. 5-vl technical report. arXiv preprint arXiv:2502.13923 (2025)
4. Bao, J., Sun, H., Deng, H., He, Y., Zhang, Z., Li, X.: Bmad: Benchmarks for medical anomaly detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 4042–4053 (2024)
5. Bergmann, P., Fauser, M., Sattlegger, D., Steger, C.: Mvtec ad–a comprehensive real-world dataset for unsupervised anomaly detection. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 9592–9600 (2019)
6. Bergmann, P., Jin, X., Sattlegger, D., Steger, C.: The mvtec 3d-ad dataset for unsupervised 3d anomaly detection and localization. arXiv preprint arXiv:2112.09045 (2021)
7. Cao, T., Zhu, J., Pang, G.: Anomaly detection under distribution shift. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 6511–6523 (2023)
8. Cao, Y., Zhang, J., Frittoli, L., Cheng, Y., Shen, W., Boracchi, G.: Adaclip: Adapting clip with hybrid learnable prompts for zero-shot anomaly detection. In: European Conference on Computer Vision. pp. 55–72. Springer (2024)
9. Chen, C., Zhu, J., Luo, X., Shen, H.T., Song, J., Gao, L.: Coin: A benchmark of continual instruction tuning for multimodel large language models. Advances in Neural Information Processing Systems **37**, 57817–57840 (2024)
10. Chen, G., Shen, L., Shao, R., Deng, X., Nie, L.: Lion: Empowering multimodal large language model with dual-level visual knowledge. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 26540–26550 (2024)
11. Chen, Z., Wu, J., Wang, W., Su, W., Chen, G., Xing, S., Zhong, M., Zhang, Q., Zhu, X., Lu, L., et al.: Internvl: Scaling up vision foundation models and aligning for generic visual-linguistic tasks. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 24185–24198 (2024)
12. Dhar, P., Singh, R.V., Peng, K.C., Wu, Z., Chellappa, R.: Learning without memorizing. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 5138–5146 (2019)
13. Dong, W., Chu, G., Pan, Z., Xie, G.S., Shan, C., Zhao, F.: Region-aware compositional context prompting for zero-shot anomaly detection. In: ECAI 2025, pp. 82–89. IOS Press (2025)
14. Feng, W., Hao, C., Zhang, Y., Han, Y., Wang, H.: Mixture-of-loras: An efficient multitask tuning for large language models. arXiv preprint arXiv:2403.03432 (2024)
15. Fu, C., Lin, H., Wang, X., Zhang, Y.F., Shen, Y., Liu, X., Cao, H., Long, Z., Gao, H., Li, K., et al.: Vita-1.5: Towards gpt-4o level real-time vision and speech interaction. arXiv preprint arXiv:2501.01957 (2025)
16. Gao, J., Li, Y., Cao, Z., Li, W.: Interleaved-modal chain-of-thought. In: Proceedings of the Computer Vision and Pattern Recognition Conference. pp. 19520–19529 (2025)
17. Gao, P., Han, J., Zhang, R., Lin, Z., Geng, S., Zhou, A., Zhang, W., Lu, P., He, C., Yue, X., et al.: Llama-adapter v2: Parameter-efficient visual instruction model. arXiv preprint arXiv:2304.15010 (2023)
18. Ge, C., Wang, X., Zhang, Z., Chen, H., Fan, J., Huang, L., Xue, H., Zhu, W.: Dynamic mixture of curriculum lora experts for continual multimodal instruction tuning. arXiv preprint arXiv:2506.11672 (2025)

19. Givnan, S., Chalmers, C., Fergus, P., Ortega-Martorell, S., Whalley, T.: Anomaly detection using autoencoder reconstruction upon industrial motors. Sensors **22**(9), 3166 (2022)
20. Guo, H., Zeng, F., Xiang, Z., Zhu, F., Wang, D.H., Zhang, X.Y., Liu, C.L.: Hidellava: Hierarchical decoupling for continual instruction tuning of multimodal large language model. arXiv preprint arXiv:2503.12941 (2025)
21. He, H., Bai, Y., Zhang, J., He, Q., Chen, H., Gan, Z., Wang, C., Li, X., Tian, G., Xie, L.: Mambaad: Exploring state space models for multi-class unsupervised anomaly detection. Advances in Neural Information Processing Systems **37**, 71162–71187 (2024)
22. Houlsby, N., Giurgiu, A., Jastrzebski, S., Morrone, B., De Laroussilhe, Q., Gesmundo, A., Attariyan, M., Gelly, S.: Parameter-efficient transfer learning for nlp. In: International conference on machine learning. pp. 2790–2799. PMLR (2019)
23. Hu, E.J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., Chen, W., et al.: Lora: Low-rank adaptation of large language models. ICLR **1**(2), 3 (2022)
24. Huai, T., Zhou, J., Wu, X., Chen, Q., Bai, Q., Zhou, Z., He, L.: Cl-moe: Enhancing multimodal large language model with dual momentum mixture-of-experts for continual visual question answering. In: Proceedings of the Computer Vision and Pattern Recognition Conference. pp. 19608–19617 (2025)
25. Huang, C., Jiang, A., Feng, J., Zhang, Y., Wang, X., Wang, Y.: Adapting visual-language models for generalizable anomaly detection in medical images. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11375–11385 (2024)
26. Jeong, J., Zou, Y., Kim, T., Zhang, D., Ravichandran, A., Dabeer, O.: Winclip: Zero-/few-shot anomaly classification and segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 19606–19616 (2023)
27. Kirkpatrick, J., Pascanu, R., Rabinowitz, N., Veness, J., Desjardins, G., Rusu, A.A., Milan, K., Quan, J., Ramalho, T., Grabska-Barwinska, A., et al.: Overcoming catastrophic forgetting in neural networks. Proceedings of the national academy of sciences **114**(13), 3521–3526 (2017)
28. Lester, B., Al-Rfou, R., Constant, N.: The power of scale for parameter-efficient prompt tuning. arXiv preprint arXiv:2104.08691 (2021)
29. Li, B., Zhang, Y., Guo, D., Zhang, R., Li, F., Zhang, H., Zhang, K., Zhang, P., Li, Y., Liu, Z., et al.: Llava-onevision: Easy visual task transfer. arXiv preprint arXiv:2408.03326 (2024)
30. Li, W., Xu, X., Gu, Y., Zheng, B., Gao, S., Wu, Y.: Towards scalable 3d anomaly detection and localization: A benchmark via 3d anomaly synthesis and a self-supervised learning network. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 22207–22216 (2024)
31. Li, X.L., Liang, P.: Prefix-tuning: Optimizing continuous prompts for generation. arXiv preprint arXiv:2101.00190 (2021)
32. Li, X., Zhang, Z., Tan, X., Chen, C., Qu, Y., Xie, Y., Ma, L.: Promptad: Learning prompts with only normal samples for few-shot anomaly detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 16838–16848 (2024)
33. Li, Y., Cao, Y., Liu, C., Xiong, Y., Dong, X., Huang, C.: Iad-r1: Reinforcing consistent reasoning in industrial anomaly detection. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 40, pp. 6583–6591 (2026)
34. Li, Z., Hoiem, D.: Learning without forgetting. IEEE transactions on pattern analysis and machine intelligence **40**(12), 2935–2947 (2017)

35. Lin, B., Ye, Y., Zhu, B., Cui, J., Ning, M., Jin, P., Yuan, L.: Video-llava: Learning united visual representation by alignment before projection. In: Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing. pp. 5971–5984 (2024)
36. Lin, J., He, Y., Xu, W., Guan, J., Zhang, J., Zhou, S.: Latent feature reconstruction for unsupervised anomaly detection. Applied Intelligence **53**(20), 23628–23640 (2023)
37. Lin, S., Yang, L., Fan, D., Zhang, J.: Beyond not-forgetting: Continual learning with backward knowledge transfer. Advances in Neural Information Processing Systems **35**, 16165–16177 (2022)
38. Liu, F., Zhu, X., Feng, P., Zeng, L.: Anomaly detection via progressive reconstruction and hierarchical feature fusion. Sensors **23**(21), 8750 (2023)
39. Liu, H., Li, C., Wu, Q., Lee, Y.J.: Visual instruction tuning. Advances in neural information processing systems **36**, 34892–34916 (2023)
40. Liu, J., Wu, K., Nie, Q., Chen, Y., Gao, B.B., Liu, Y., Wang, J., Wang, C., Zheng, F.: Unsupervised continual anomaly detection with contrastively-learned prompt. In: Proceedings of the AAAI conference on artificial intelligence. vol. 38, pp. 3639–3647 (2024)
41. Liu, J., Xie, G., Chen, R., Li, X., Wang, J., Liu, Y., Wang, C., Zheng, F.: Real3d-ad: A dataset of point cloud anomaly detection. Advances in Neural Information Processing Systems **36**, 30402–30415 (2023)
42. Liu, Z., Zhou, Y., Xu, Y., Wang, Z.: Simplenet: A simple network for image anomaly detection and localization. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 20402–20411 (2023)
43. Lopez-Paz, D., Ranzato, M.: Gradient episodic memory for continual learning. Advances in neural information processing systems **30** (2017)
44. Loshchilov, I., Hutter, F.: Decoupled weight decay regularization. arXiv preprint arXiv:1711.05101 (2017)
45. Piao, H., Wu, Y., Wu, D., Wei, Y.: Federated continual learning via prompt-based dual knowledge transfer. In: Forty-first International Conference on Machine Learning (2024)
46. Qiao, F., Mahdavi, M.: Learn more, but bother less: parameter efficient continual learning. Advances in Neural Information Processing Systems **37**, 97476–97498 (2024)
47. Rolnick, D., Ahuja, A., Schwarz, J., Lillicrap, T., Wayne, G.: Experience replay for continual learning. Advances in neural information processing systems **32** (2019)
48. Roth, K., Pemula, L., Zepeda, J., Schölkopf, B., Brox, T., Gehler, P.: Towards total recall in industrial anomaly detection. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 14318–14328 (2022)
49. Tang, J., Lu, H., Xu, X., Wu, R., Hu, S., Zhang, T., Cheng, T.W., Ge, M., Chen, Y.C., Tsung, F.: An incremental unified framework for small defect inspection. In: European conference on computer vision. pp. 307–324. Springer (2024)
50. Tao, F., Xie, G.S., Zhao, F., Shu, X.: Kernel-aware graph prompt learning for few-shot anomaly detection. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 39, pp. 7347–7355 (2025)
51. Tien, T.D., Nguyen, A.T., Tran, N.H., Huy, T.D., Duong, S., Nguyen, C.D.T., Truong, S.Q.: Revisiting reverse distillation for anomaly detection. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 24511–24520 (2023)

52. Touvron, H., Lavril, T., Izacard, G., Martinet, X., Lachaux, M.A., Lacroix, T., Rozière, B., Goyal, N., Hambro, E., Azhar, F., et al.: Llama: Open and efficient foundation language models. arXiv preprint arXiv:2302.13971 (2023)
53. Wang, G., Han, S., Ding, E., Huang, D.: Student-teacher feature pyramid matching for anomaly detection. arXiv preprint arXiv:2103.04257 (2021)
54. Wang, H., Yan, W., Lai, Q., Geng, X.: Temporal consistency-aware text-to-motion generation. Visual Intelligence **4**(1), 7 (2026)
55. Wang, X., Chen, T., Ge, Q., Xia, H., Bao, R., Zheng, R., Zhang, Q., Gui, T., Huang, X.J.: Orthogonal subspace learning for language model continual learning. In: Findings of the Association for Computational Linguistics: EMNLP 2023. pp. 10658–10671 (2023)
56. Wang, Z., Che, C., Wang, Q., Li, Y., Shi, Z., Wang, M.: Smolora: Exploring and defying dual catastrophic forgetting in continual visual instruction tuning. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 177–186 (2025)
57. Wei, Q., Ren, Y., Hou, R., Shi, B., Lo, J.Y., Carin, L.: Anomaly detection for medical images based on a one-class classification. In: Medical Imaging 2018: Computer-Aided Diagnosis. vol. 10575, pp. 375–380. SPIE (2018)
58. Wu, J., Zhong, M., Xing, S., Lai, Z., Liu, Z., Chen, Z., Wang, W., Zhu, X., Lu, L., Lu, T., et al.: Visionllm v2: An end-to-end generalist multimodal large language model for hundreds of vision-language tasks. Advances in Neural Information Processing Systems **37**, 69925–69975 (2024)
59. Xu, J., Lo, S.Y., Safaei, B., Patel, V.M., Dwivedi, I.: Towards zero-shot anomaly detection and reasoning with multimodal large language models. In: Proceedings of the Computer Vision and Pattern Recognition Conference. pp. 20370–20382 (2025)
60. Zhang, X., Xu, M., Qiu, D., Yan, R., Lang, N., Zhou, X.: Mediclip: Adapting clip for few-shot medical image anomaly detection. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 458–468. Springer (2024)
61. Zhao, H., Li, Y., He, N., Ma, K., Fang, L., Li, H., Zheng, Y.: Anomaly detection for medical images using self-supervised and translation-consistent features. IEEE Transactions on Medical Imaging **40**(12), 3641–3651 (2021)
62. Zhao, H., Zhu, F., Guo, H., Wang, M., Wang, R., Meng, G., Zhang, Z.: Mllm-cl: Continual learning for multimodal large language models. arXiv preprint arXiv:2506.05453 (2025)
63. Zheng, H., Xu, T., Sun, H., Pu, S., Chen, R., Sun, L.: Thinking before looking: Improving multimodal llm reasoning via mitigating visual hallucination. arXiv preprint arXiv:2411.12591 (2024)
64. Zhou, Q., Pang, G., Tian, Y., He, S., Chen, J.: Anomalyclip: Object-agnostic prompt learning for zero-shot anomaly detection. arXiv preprint arXiv:2310.18961 (2023)
65. Zhu, F., Zhang, X.Y., Wang, C., Yin, F., Liu, C.L.: Prototype augmentation and self-supervision for incremental learning. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 5871–5880 (2021)