Expert Systems With Applications 330 (2026) 133055

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Continual test-time adaptation for object detection with adaptive monitoring and randomized restoration

Shilei Cao [a], Juepeng Zheng [a,b,*], Yan Liu [c], Baoquan Zhao [a], Ziqi Yuan [d], Weijia Li [e], Runmin Dong [a], Haohuan Fu [e,b,f]

[a] School of Artificial Intelligence, Sun Yat-Sen University, Zhuhai, 519080, China
[b] National Supercomputing Center in Shenzhen, Shenzhen, 518055, China
[c] School of Information Science and Technology, University of Science and Technology of China, Hefei, 230026, China
[d] State Key Laboratory of Intelligent Technology and Systems, Department of Computer Science and Technology, Tsinghua University, Beijing, 100084, China
[e] Tsinghua Shenzhen International Graduate School, Tsinghua University, Shenzhen, 518071, China
[f] Ministry of Education Key Laboratory for Earth System Modeling and the Department of Earth System Science, Tsinghua University, Beijing, 100084, China

ARTICLE INFO

Keywords:
Unsupervised domain adaptation
Test-time adaptation
Continual test-time adaptation
Object detection

ABSTRACT

Real-world application models are commonly deployed in dynamic environments, where the target domain distribution undergoes temporal changes. Continual Test-Time Adaptation (CTTA) has recently emerged as a promising technique to gradually adapt a source-trained model to continually changing target domains. Despite recent advancements in addressing CTTA, two critical issues remain: *1)* Fixed thresholds for pseudo-labeling in existing methodologies lead to low-quality pseudo-labels, as model confidence varies across categories and domains; *2)* Stochastic parameter restoration methods for mitigating catastrophic forgetting fail to preserve critical information effectively, due to their intrinsic randomness. To tackle these challenges for detection models in CTTA scenarios, we present AMROD, featuring three core components. Firstly, the object-level contrastive learning module extracts object-level features for contrastive learning to refine the feature representation in the target domain. Secondly, the adaptive monitoring module dynamically skips unnecessary adaptation and updates the category-specific threshold based on predicted confidence scores to enable efficiency and improve the quality of pseudo-labels. Lastly, the adaptive randomized restoration mechanism selectively reset inactive parameters with higher possibilities, ensuring the retention of essential knowledge. We demonstrate the effectiveness of AMROD on four CTTA object detection tasks, where AMROD outperforms existing methods, especially achieving a 3.2 mAP improvement and a 20% increase in efficiency on the Cityscapes-to-Cityscapes-C CTTA task. The code of this work is available at https://github.com/ShileiCao/AMROD.

1. Introduction

Deep learning models have demonstrated immense potential across various vision tasks such as image recognition (He et al., 2016), object detection (Ren et al., 2015), and image segmentation (Long et al., 2015). However, these models experience pronounced degradation in performance when confronted with training data (i.e., source domain) and testing data (i.e., target domain) originating from disparate distributions. This phenomenon, commonly referred to as distribution shifts, poses a significant challenge (Hendrycks & Dietterich, 2019). In such a scenario, unsupervised domain adaptation (UDA) becomes crucial, which typically involves aligning the distributions of source and target data, thereby mitigating the impact of distribution shifts (Jiang et al., 2025; Zhang et al., 2022). Still, UDA falls short by necessitating access to source data, which is often inaccessible due to privacy constraints, proprietary data concerns, or data transmission barriers (VS et al., 2023b). This limitation catalyzes the exploration of Test-Time Adaptation (TTA) where the source-trained model directly adapts toward unlabeled test samples encountered during evaluation in an online manner, without the reliance on the source data (Wang et al., 2020). Nonetheless, the TTA methods, which assume a static target domain, face a more challenging and realistic problem, as real-world systems work in non-stationary environments. For example, a vehicle may encounter various continuous environmental changes such as fog, night, rain, and snow during its

* Corresponding author.
E-mail address: zhengjp8@mail.sysu.edu.cn (J. Zheng).

https://doi.org/10.1016/j.eswa.2026.133055
Received 3 October 2025; Received in revised form 2 April 2026; Accepted 27 May 2026
Available online 29 May 2026

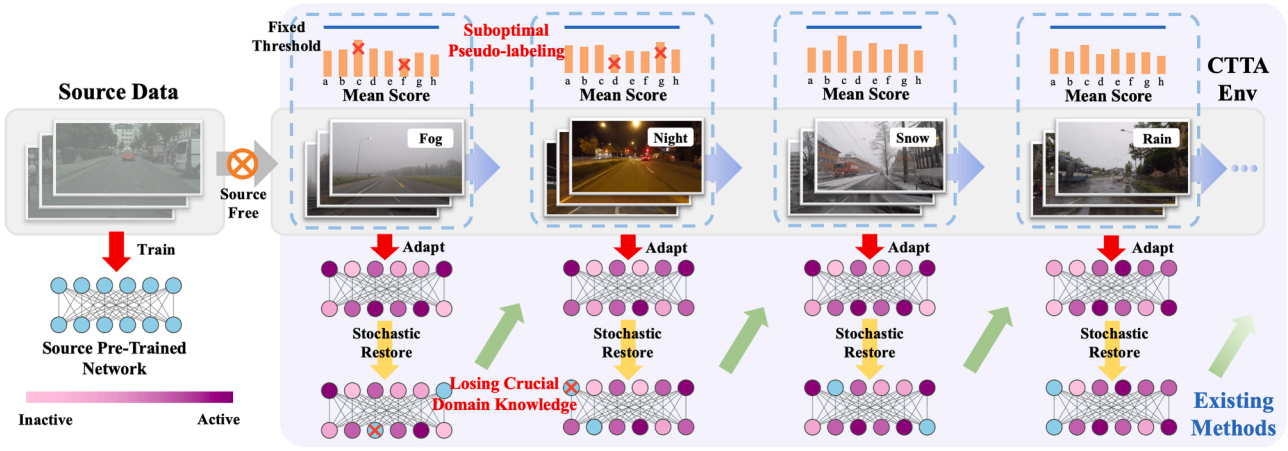

Fig. 1. Motivation. (1) Top: Model confidence fluctuates across different categories and target domains. Applying a uniformly fixed threshold fails to capture these variations, leading to suboptimal pseudo-labeling. (2) Bottom: Stochastic restoration randomly resets neurons to mitigate forgetting (darker colors indicate higher activity, while blue represents source knowledge). This randomness can inadvertently erase crucial, domain-specific knowledge acquired from adaptation.

journey. Existing TTA methods are vulnerable to catastrophic forgetting of source knowledge and error accumulation when adaptation faces more than one distribution shift (Wang et al., 2022). These issues become more pronounced when adaptation occurs in environments where the target domain is not only dynamic but also evolves over time.

Recently, Wang et al. (2022) introduce CoTTA by applying stochastic parameters restoration to mitigate catastrophic forgetting in Continual Test-Time Adaptation (CTTA) scenarios, where the model is continually adapted to sequences of target domains. Although randomly resetting parameters partially helps alleviate the forgetting of source knowledge (Wang et al., 2022), its randomness may also contribute to losing crucial knowledge specific to the current domain (refer to Fig. 1). Additionally, existing CTTA methods normally utilize pseudo-labeling through a fixed threshold for self-supervision (Döbler et al., 2023; Wang et al., 2022). Given that model confidence may vary across categories and domains (refer to Fig. 1), employing a uniformly fixed threshold could exclude high-quality pseudo-labels while incorporating incorrect ones. Furthermore, these methods adapt to every new incoming data, resulting in computational inefficiency and potential performance degradation. Low-quality pseudo-labels and unstable adaptation to noisy data lead to error accumulation, as they provide negative feedback to the model, undermining its performance.

Therefore, two significant challenges persist in CTTA: Current self-training-based methods suffer from low-quality pseudo-labels, leading to error accumulation (***Challenge 1***); Continual adaptation to dynamic environments struggles to effectively retain valuable knowledge about current domain while mitigating forgetting of source knowledge (***Challenge 2***).

To tackle these challenges, we propose AMROD (**A**daptive **M**onitoring and **R**estoration for **O**bject **D**etection). Aligning with previous CTTA works for robust adaptation (Döbler et al., 2023; Wang et al., 2022), AMROD is constructed upon the mean-teacher framework (Tarvainen & Valpola, 2017). This framework involves a student model supervised by a teacher model, where the teacher model is an exponential moving average of the student model. Particularly, AMROD comprises Object-level Contrastive Learning (OCL), Adaptive Monitoring (AM), and Adaptive Randomized Restoration (ARR) modules.

More specifically, OCL extracts object-level features based on Region Proposal Network (RPN) proposals, which provide multiple cropped views around the object at different locations and scales. Subsequently, Contrastive Learning (CL) loss is applied on the proposals to guide the model to encourage similar object instances to remain close while pushing dissimilar ones apart. The OCL is well integrated into the mean-teacher paradigm as a drop-in enhancement to acquire more fine-grained feature representation. Furthermore, to make adaptation more efficient and stable and to improve the quality of pseudo-labels, i.e., addressing ***Challenge 1***, we design an AM module to decide whether to pause or resume the adaptation and dynamically adjust the category-specific thresholds, based on the mean predicted confidence scores. The dynamic nature of the AM method makes it better suited to address the effects of continuously changing distributions. Finally, the proposed ARR mechanism resets inactive parameters with a higher possibility than active ones, by utilizing Fisher information (Fisher, 1922) as an indicator of parameter importance while incorporating a stochastic component. ARR not only helps prevent forgetting but preserves important information, i.e., addressing ***Challenge 2***. On the other hand, its randomness allows falsely activated parameters to be reset as well, thereby resulting in greater stability for adaptation.

We demonstrate the effectiveness of AMROD on four CTTA object detection benchmarks, which involve synthetic and real-world distribution shifts in the short- and long-term adaptation, i.e. Cityscapes (Cordts et al., 2016), Cityscapes-C (Hendrycks & Dietterich, 2019), SHIFT (Sun et al., 2022), and ACDC (Sakaridis et al., 2021) datasets. Results indicate that our method significantly improves performance over existing methods, with gains of up to 3.2 mAP and 20% in computational efficiency. Our main contributions are:

- This study introduces AMROD, which pioneers in exploring CTTA for detection models. Specifically, we propose to leverage object-level features for contrastive learning to refine feature representation in CTTA object detection, bypassing the computational burden of large batch sizes typically required for contrastive learning.
- To address the two challenges in CTTA, our proposed AM module enables dynamic skipping and category-specific threshold updates based on the mean predicted scores, therefore improving robustness and pseudo-label quality. Moreover, the ARR module resets the inactive parameter with higher possibilities, effectively preventing error accumulation and catastrophic forgetting.
- Empirical experiments demonstrate that AMROD surpasses existing methods and facilitates short-term and long-term adaptation under both synthetic and real-world continual distribution shift, notably achieving up to a 3.2 mAP performance gain and a 20% increase in computational efficiency on the Cityscapes-to-Cityscapes-C task.

2. Related works

2.1. Source-free domain adaptation

UDA tackles the inter-domain divergence by aligning the distributions of source and target data (Cao et al., 2025; Li et al., 2022b, 2025; Liang et al., 2025a,b; Saito et al., 2019, 2018; Xu et al., 2025). Despite its effectiveness, the limitation of UDA lies in its requirement for access to the source domain data, which often raises concerns regarding data privacy and transmission efficiency (Huang et al., 2021; VS et al., 2023b). As a result, Source-Free Domain Adaptation (SFDA) received extensive research attention, where the source-trained detector is adapted to the target training data without any source data (Chen et al., 2023; Li et al., 2022a; Lü et al., 2024; VS et al., 2023a,b; Wang et al., 2025; Ye et al., 2025) before evaluation. For instance, MemCLR (VS et al., 2023a) employs a cross-attention-based memory bank with CL for source-free detectors, while IRG (VS et al., 2023a) utilizes the object relations with instance relation graph network to explore the SFDA setting for object detection. However, the standard SFDA setting requires prior knowledge of the target domain, which is impractical in most real-world applications.

2.2. Test-time adaptation

TTA adapts the source-trained model to the target test data during inference time without access to the source data. Since both TTA and SFDA involve adapting the source-trained model to the unlabeled target data without utilizing source data, some works also refer to TTA as SFDA (Brahma & Rai, 2023; Wang et al., 2022). In this paper, we distinguish between TTA and SFDA based on evaluation protocol, although they can be transformed into each other in experiments. Furthermore, TTA methods improve the model performance under distribution shift commonly through pseudo-labeling (Iwasawa & Matsuo, 2021; Liang et al., 2025b; Sun et al., 2020; Zeng et al., 2023), batchnorm statistics updating (Hu et al., 2021; You et al., 2021), or entropy regularization (Fu et al., 2025; Iwasawa & Matsuo, 2021; Liang et al., 2025c; Niu et al., 2022; Wang et al., 2020) during testing. For example, Tent (Wang et al., 2020) updates the batchnorm parameters with entropy minimization and demands a large batch size for optimization during test-time adaptation, which is unsuitable for real-time detection model deployment where images are processed sequentially. The above approaches assume a static target domain where the target data come from a single domain. However, in practical scenarios, the distribution of target domain may exhibit a continual shift over time.

2.3. Continual test-time adaptation

Conceptually, the core principles of CTTA are closely related to the classical paradigm of on-line retrainable neural networks (An et al., 2011; Doulamis et al., 2000; Ioannou et al., 2006; Kollias et al., 2016). Originally explored in dynamic vision and signal processing tasks such as video segmentation and continuous image analysis (Doulamis et al., 2000; Ntalianis et al., 2002), retrainable neural networks are designed to continuously update their parameters during the operational (inference) phase to adapt to non-stationary environments. A critical challenge in these early dynamic systems, as in modern CTTA, is the forgetting issues. To prevent the loss of previously acquired knowledge while adapting to new conditions, these historical structures explicitly handle the forgetting issue through constrained optimization and memory management. For instance, systems employ retraining algorithms that explicitly minimize weight modifications relative to previous model states (Doulamis et al., 2000, 2002) and utilize selective retraining frameworks that actively maintain representative historical samples alongside new observations (An et al., 2011).

Modern CTTA formalizes this continuous adaptation challenge under unsupervised constraints. Early modern works consider adaptation to evolving and continually changing domains by aligning the source and target data (Hoffman et al., 2014). These methods rely on source data during inference, which limits their applicability. Recently, Wang et al. (2022) introduce CoTTA, marking the first work tailored to the demands of CTTA by adapting a pre-trained model to sequences of domains without source data. Subsequently, research efforts have been dedicated to exploring CTTA, primarily focused on classification (Brahma & Rai, 2023; Döbler et al., 2023; Gan et al., 2023a; Niloy et al., 2024; Yu et al., 2023) and segmentation (Liu et al., 2023; Ni et al., 2023; Niloy et al., 2024; Song et al., 2023; Zhu et al., 2023) tasks. For instance, similar to CoTTA, Zhu et al. (2023) apply a stochastic reset mechanism to prevent forgetting in the CTTA medical segmentation task. In contrast, PETAL (Brahma & Rai, 2023) utilizes the Fisher Information Matrix (FIM) as a metric of parameter importance to reset only the most irrelevant parameters across all layers. Nevertheless, these two methods are suboptimal, as the randomness might lead to losing essential information, and pure data-driven restoration may retain false active parameters resulting from noise.

Moreover, the mean-teacher framework (Tarvainen & Valpola, 2017) serves as a base architecture for most CTTA works (Döbler et al., 2023; Gong et al., 2022; Niloy et al., 2024; Wang et al., 2022), where the teacher model generates pseudo-labels via a fixed threshold to supervise the training of the student model. Nonetheless, these methods suffer from low-quality pseudo-labels with a uniform threshold since the model confidence varies across categories and domains. Furthermore, Wang et al. (2024) design a dynamic thresholding technique to update the threshold in a batch for classification tasks, requiring a large batch size. However, this approach is unsuitable for object detection where smaller batch sizes are preferred for computational efficiency. For object detection, Gan et al. (2023b) present a cloud-device collaborative continual adaptation paradigm by aligning source and target distribution. More recently, Yoo et al. (2024) explore primarily on short-term CTTA by utilizing feature distribution statistics calculated from source samples to determine whether to update a lightweight adapter integrated into the backbone. Nevertheless, the reliance on accessing source data in both methods restricts their practical applicability due to privacy and resource limitations. In contrast, AMROD operates under a strictly source-free constraint without relying on any source data or architectural additions, addressing forgetting for both short-term and long-term CTTA. Therefore, a gap still exists in exploring the CTTA in object detection to improve the quality of pseudo-label (Challenge 1) and effectively reset noisy neurons (Challenge 2), without relying on source data.

2.4. Continual learning

Continual learning, also known as incremental learning, typically involves enabling the model to retain previously acquired knowledge while learning from sequences of tasks (De Lange et al., 2021). It is commonly categorized into replay methods (Rebuffi et al., 2017; Tiwari et al., 2022), parameter isolation method (Aljundi et al., 2017; Xu & Zhu, 2018), and regularization-based methods (Kirkpatrick et al., 2017; Li & Hoiem, 2017). As an illustration, Elastic weight consolidation (EWC) (Kirkpatrick et al., 2017) is a regularization-based technique that penalizes the changing of parameters with a significant impact on prediction, based on the Fisher Information Matrix (FIM). In this paper, motivated by Brahma and Rai (2023), Kirkpatrick et al. (2017), we adopt the FIM as a metric of parameter importance for resetting noisy parameters. Furthermore, we introduce randomness to enhance model robustness during continual adaptation. Additionally, while the continual learning approaches aim to tackle catastrophic forgetting in sequences of new tasks, our work focuses on learning from different domains for a single task.

Table 1
Comparisons between different problem settings.

Setting	Source Data	Target Training Data	Target Distribution	Train Loss	Test Loss	Online
Continual Learning	×	(x^t, y^t)	Dynamic	$\mathcal{L}(x^t, y^t)$	×	×
Unsupervised Domain Adaptation	(x^s, y^s)	(x^t)	Static	$\mathcal{L}(x^s, y^s) + \mathcal{L}(x^s, x^t)$	×	×
Source-free Domain Adaptation	×	(x^t)	Static	$\mathcal{L}(x^t)$	×	×
Test-time Adaptation	×	×	Static	×	$\mathcal{L}(x^t)$	✓
Continual Test-Time Adaptation	×	×	Dynamic	×	$\mathcal{L}(x^t)$	✓

2.5. Knowledge distillation

The concept of Knowledge Distillation (KD) originally emerged as a model compression technique designed to transfer learned knowledge from a cumbersome, complex teacher model into a more compact student model (Buciluǎ et al., 2006; Hinton et al., 2015). In the realm of object detection, KD has been adapted to handle complex localization and regression tasks, effectively transferring fine-grained, region-based feature representations and relational knowledge between domains (Chen et al., 2017; Tian et al., 2021). Beyond model compression, recent studies demonstrate that distilling knowledge from an adapted or stabilized teacher network can effectively mitigate catastrophic forgetting in exemplar-free continual learning scenarios (Szatkowski et al., 2024) and facilitate domain-aware continual generalization (Reddy et al., 2024).

This concept is highly relevant to our proposed AMROD and most CTTA frameworks, which utilize a mean-teacher architecture (Tarvainen & Valpola, 2017). The mean-teacher paradigm functions as an online, self-knowledge distillation mechanism, which relies on the slowly updating teacher model to distill historically stabilized knowledge into the dynamically updating student model via pseudo-labels. This distillation process acts as an anchor against continuous domain shifts, complementing our ARR module to ensure that the network acquires new target-domain representations without catastrophically forgetting its foundational object detection capabilities.

2.6. Position of the proposed work

While existing methods tackle isolated aspects of domain adaptation, they fall short in fully source-free, continually changing environments for object detection. Unlike standard UDA or SFDA methods that assume a static target domain, AMROD is explicitly designed to handle continual, non-stationary distribution shifts. Furthermore, unlike modern CTTA approaches for object detection that still rely heavily on source data, AMROD positions itself as a strictly standalone, fully source-free framework.

AMROD effectively bridges the gaps between TTA, Continual Learning, and Knowledge Distillation. By utilizing the mean-teacher paradigm as an online self-knowledge distillation anchor, AMROD stabilizes the adaptation process. It directly addresses the catastrophic forgetting inherent in CTTA through the ARR module, while simultaneously overcoming the noisy pseudo-labeling problem of traditional TTA through the AM model. Ultimately, this positions AMROD as a highly efficient, robust solution uniquely tailored for the complex demands of real-world continual object detection.

3. Method

3.1. Preliminary

3.1.1. Problem statement

Given a sequence of domain $D = \{d_i\}_{i=0}^{n}$, we define d_0 as the source domain and the subsequent domains as the target domain. The objective of CTTA is to enhance the performance of the model $M_{\theta^0}(x)$, where the parameters θ^0 are pre-trained on source data x_{d_0}, y_{d_0} from d_0, in a continually changing target domain during inference time without using source data. For simplicity, we hereafter denote the data without the subscript about the specific domain. At time step t, unlabeled target data x^t is provided sequentially, following the domain group order. The model is required to make a prediction $y^t = M_{\theta^{t-1}}(x^t)$ using the parameters θ^{t-1} which have been updated based on previous target data $x^1, \ldots, x^{t-1}$. Subsequently, y^t serves as the evaluation output at time step t, and the model will adapt itself toward x^t as θ^t, which will only influence future inputs x^{t+n}.

Moreover, we compare CTTA with other settings in Table 1. These settings are developed to meet the diverse prerequisites and requirements for real-world applications.

3.1.2. Mean-teacher framework

Following previous CTTA works (Wang et al., 2022), we build AMROD based on the mean-teacher framework (Tarvainen & Valpola, 2017), which features the interplay between a teacher model and a student model. Specifically, both networks are first initialized with the source-trained model. The teacher model produces the pseudo-labels $\hat{y}^t$ based on teacher prediction y^t for the unlabeled target data to supervise the student model. While the parameters of the student are optimized via gradient descent, the teacher is updated following an Exponential Moving Average (EMA) strategy based on the student. Formally, this process can be expressed as follows:

$$\mathcal{L}_{pl}(x^t) = \mathcal{L}_{rpn}(x^t, \hat{y}^t) + \mathcal{L}_{rcnn}(x^t, \hat{y}^t), \tag{1}$$

$$\theta_S^t \leftarrow \theta_S^{t-1} + \gamma \frac{\partial(\mathcal{L}_{pl}(x^t))}{\partial \theta_S^{t-1}}, \tag{2}$$

$$\theta_T^t \leftarrow \alpha\theta_T^{t-1} + (1-\alpha)\theta_S^t, \tag{3}$$

where x^t and $\hat{y}^t$ denote the unlabeled target data and corresponding pseudo-labels at time step t. θ_S^t and θ_T^t symbolize the parameters of student and teacher networks. Moreover, the supervision loss $\mathcal{L}_{pl}$ in Faster-RCNN (Ren et al., 2015), which consists of the RPN loss $\mathcal{L}_{rpn}$ and RCNN loss $\mathcal{L}_{rcnn}$, is utilized for pseudo-labeling. Additionally, γ represents the student's learning rate, and the EMA rate is denoted by α. Hence, the teacher can be regarded as an ensemble of historical students, providing stable supervision. However, this framework encounters error accumulation and catastrophic forgetting in dynamic environments. Therefore, we design AMROD to fit in the CTTA setting.

3.2. Proposed method

3.2.1. Overview

Fig. 2 presents an overview of AMROD with details described in Algorithm 1, consisting of Object-level Contrastive Learning (OCL), Adaptive Monitoring (AM), and Adaptive Randomized Restoration (ARR). Inspired by UDA object detection methods (Li et al., 2022b; Liu et al., 2021), we adopt the Weak-Strong augmentation to enable the teacher model to generate reliable pseudo-labels without being affected by heavy augmentation. Specifically, the teacher receives input with weak augmentations, while we input student networks with strong enhanced images. The overall loss of the student is defined as:

$$\mathcal{L}_{all} = \mathcal{L}_{pl}(x^t) + \lambda\mathcal{L}_{cl}(x^t) + \mu\mathcal{L}_{kl}(x^t), \tag{4}$$

where $\mathcal{L}_{cl}(x^t)$ denotes the contrastive learning (CL) loss, to be elaborated later. The $\mathcal{L}_{kl}$ represents the Kullback-Leibler (KL) Divergence (Kullback & Leibler, 1951) loss, used to quantify the distinction between

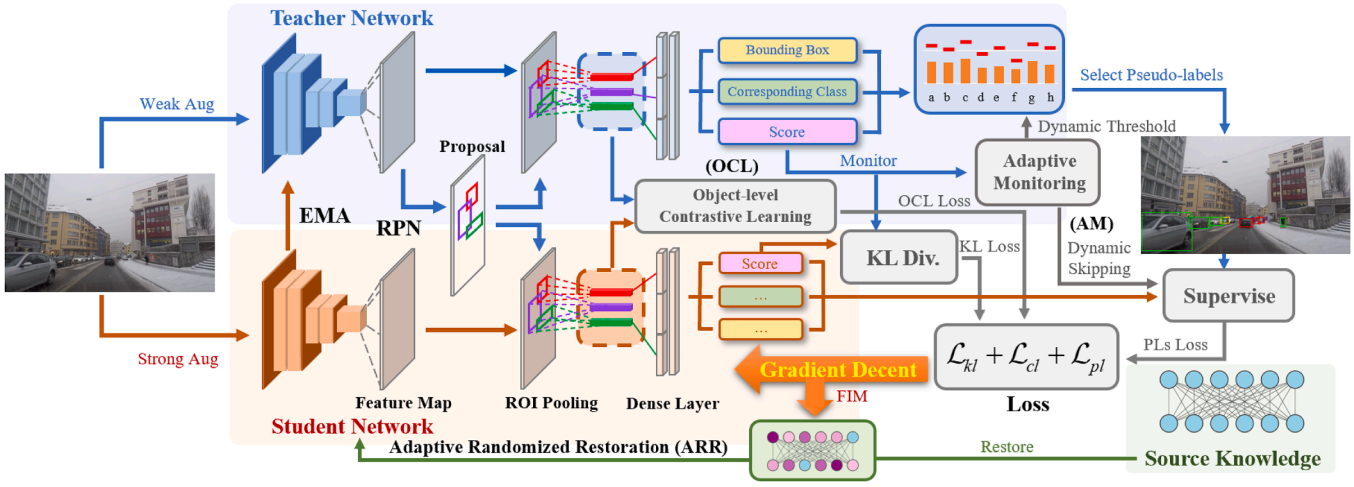

Fig. 2. The overview of the proposed method. AMROD follows the mean-teacher framework, featuring: *(1)* The OCL module compares region of interest features extracted from feature maps of both networks based on teacher proposals for contrastive learning; *(2)* The AM module dynamically skips unnecessary adaptation and adjusts category-specific thresholds based on the mean prediction scores; *(3)* The ARR strategy reset the inactive parameter with higher possibilities based on the FIM.

two probability distributions. The parameters λ and μ are corresponding weights. KL loss is defined as:

$$\mathcal{L}_{kl}(P \parallel Q) = \sum_{x \in \mathcal{X}} P(x) \log\left(\frac{P(x)}{Q(x)}\right), \tag{5}$$

where $P(x)$ and $Q(x)$ represent different distribution. We employ the KL divergence loss to encourage the student model to approximate the teacher model closely.

3.2.2. Object-level contrastive learning

SimCLR (Chen et al., 2020) is a widely used CL approach for self-supervised learning, which learns high-quality feature representation across differently augmented views of the same image. Notably, the SimCLR is initially designed for classification tasks, assuming each image pertains to a single category, and requires large batch sizes to ensure sufficient positive and negative pairs for representation learning. Consequently, the original SimCLR is not well-suited for object detection, where images typically contain multiple instances, thus requiring significant computational resources to accommodate large batch sizes.

Motivated by SimCLR (Chen et al., 2020), we present an OCL module to extract teacher and student features for CL based on proposals from the RPN. This strategy provides multiple cropped views around the object instance, eliminating the need for large batch sizes and ensuring computational efficiency for online detector updates. Specifically, given a weakly augmented image $A_{weak}(x^t)$ at time step t, the teacher produces l ROI proposals $P^t = \{p_i^t\}_{i=1}^{l}$ via region proposal network. OCL then apply RoIAlign (He et al., 2017) to extract corresponding teacher and student object-level features $T^t = \{t_i^t \in \mathbb{R}^{1\times C}\}_{i=1}^{l}$ and $S^t = \{s_i^T \in \mathbb{R}^{1\times C}\}_{i=1}^{l}$ based on the feature map from the backbone, respectively. The features associated with the same proposal are considered positive pairs, otherwise negative pairs. Then CL loss is applied to these features T^t and S^t images by minimizing:

$$\mathcal{L}_{cl}(x^t) = \frac{1}{l}\sum_{i=1}^{l} -\log \frac{\exp(t_i^t \cdot s_i^t/\tau)}{\sum_{j=1}^{l}\exp(t_i^t \cdot s_j^t/\tau)}, \tag{6}$$

where $\tau > 0$ is the temperature, and t_i^t and s_i^t denote the features of two different augmentations of the same object, serving as the positive pair. This strategy encourages the model to learn fine-grained and localized feature representations on the target domain, without relying on accurate pseudo-labels. Moreover, the OCL is well integrated into the mean-teacher self-training paradigm as a drop-in enhancement for feature adaptation.

Algorithm 1 Pseudo-code for the AMROD.

Input: Unlabeled test data x^t for the target domain
for each iteration t **do**
  Generate predictions y^t using the teacher model
  // 1. Adaptive Monitoring
  Compute the index $\frac{\vec{l}^t}{\vec{l}_{ema}^{t-1}}$ based on teacher prediction
  Udapate the moving average $\vec{l}_{ema}^{t}$ through Eq. (7)
  if $\frac{\vec{l}^t}{\vec{l}_{ema}^{t-1}}$ not lies with $(\frac{1}{\delta_s}, \delta_s)$ **then**
    Pause the current adaption
  end if
  for each category c **do**
    Compute mean predicted scores $\overline{l_c^t}$ of category c
    Update dynamic thresholds through Eq. (8)
  end for
  // 2. Object-level Contrastive Learning
  Extract teacher features T^t and student features S^t based on teacher proposal
  Compute contrastive learning loss through Eq. (1)
  Generate the pseudo-label $\hat{y}^t$ through δ_c^t
  Update the student model θ_S^t through Eq. (2) with the supervised loss through Eq. (6) and teacher model θ_T^t through Eq. (3)
  // 3. Adaptive Randomized Restoration
  Generate random matrix $R^t \sim \text{Uniform}(0, 1)$
  Generate the FIM F^t through Eq. (12)
  Generate reset score matrix W^t through Eq. (13)
  Find the q-quantile of W^t: $\eta = quantile(F^t, q)$
  Generate the mask matrix M^t through Eq. (14)
  Reset the updated student model through Eq. (15)
end for
Output: Teacher predictions y^t

3.2.3. Adaptive monitoring

The AM module monitors the model's status using the predicted scores with dynamic skipping and dynamic thresholds. The former halts unnecessary adaptations to save computational resources, while the latter updates the category-specific threshold dynamically.

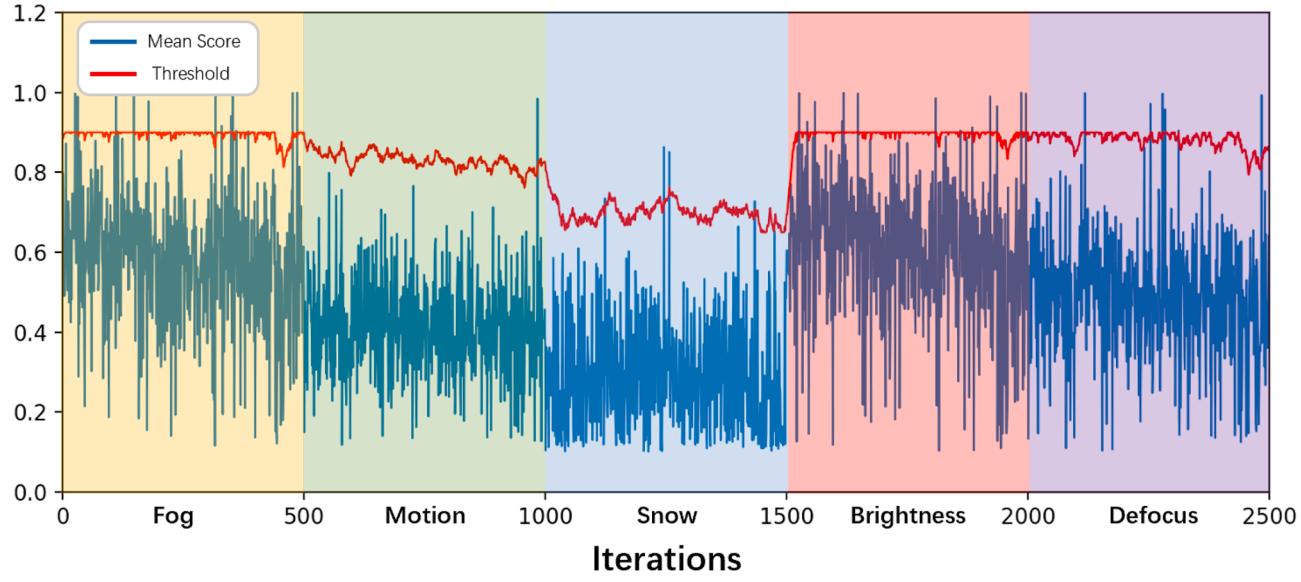

Fig. 3. Visualization of the mean score and the dynamic threshold for the "car" category during adaptation across five different domains from the Cityscapes-C dataset. The threshold (red) dynamically adjusts to the model's mean confidence score (blue) as the environment changes.

Specifically, the teacher first make predictions $y^t = M_{\theta_T^{t-1}}(A_{weak}(x^t))$ for the weakly enhanced images. AM then computes the overall mean scores across all predicted instances $\vec{l}^t$ in teacher's predictions and maintains its exponentially moving average $\vec{l}^t_{ema}$ through:

$$\vec{l}^t_{ema} \leftarrow \beta_s \cdot \vec{l}^{t-1}_{ema} + (1-\beta_s)\cdot \vec{l}^t, \tag{7}$$

where β_s represents the update rate. Adaptation resumes when the $\frac{\vec{l}^t}{\vec{l}^{t-1}_{ema}}$ lies within the range $(\frac{1}{\delta_s}, \delta_s)$, indicating stability, and pauses otherwise, suggesting changes in data characteristics. This dynamic skipping strategy effectively prevents unstable and noisy adaptation, thus conserving computational resources and mitigating error accumulation.

For dynamic threshold, the thresholds for each category are initialized with the same value δ^0 at first, which will be updated every iteration by:

$$\delta^t_c \leftarrow \beta_t \cdot \delta^{t-1}_c + (1-\beta_t)\cdot \epsilon \cdot (\vec{l}^t_c)^{\frac{1}{2}}, \tag{8}$$

where δ^t_c denotes the threshold of category c at time step t, $\vec{l}^t_c$ is the mean predicted scores of the category c at time step t, β_t represents the update rate of dynamic threshold, and ϵ provides a linear projection. Furthermore, the threshold δ^t_c will not change if class c does not exist in prediction $\hat{y}^t$, and we set a fixed upper and lower bound $\delta_{\max}$ and δ_{mini}. For unlabeled data from dynamic environments, this mechanism effectively generates an appropriate threshold for each category. Furthermore, Fig. 3 visualizes this process for the "car" category to provide intuition. As the environment shifts from "motion" to "snow", the model's average confidence (blue line) drops significantly. The dynamic threshold (red line) adaptively follows this trend, lowering itself to continue accepting high-quality pseudo-labels relative to the new, more challenging domain. Conversely, when the environment changes to "Brightness", the model's confidence increases, and the threshold rises accordingly.

3.2.4. Adaptive randomized restoration

Long-term adaptation to dynamic scenarios can lead to catastrophic forgetting, where self-training may reinforce erroneous predictions. Even worse, the model might fail to recover when transitioning to a new domain. Existing methods address catastrophic forgetting through stochastic reset (Wang et al., 2022) or pure data-driven restoration (Brahma & Rai, 2023). Nevertheless, the former may reset valuable parameters, potentially erasing essential knowledge relevant to the current domain. On the other hand, the latter restores the least important parameters, which may retain noise parameters, thereby leading to error accumulation. Moreover, PETAL (Brahma & Rai, 2023), designed for classification tasks, resets the least active parameters globally across the entire network. This global strategy can overlook the non-uniform scales of weights in different layers, particularly in deeper neural networks. Our approach refines this for object detection by identifying inactive parameters within each layer individually. This layer-specific technique accounts for varying parameter scales at different model depths. Building upon this, our proposed ARR mechanism further incorporates randomness to reset irrelevant parameters with a higher probability, thereby enhancing stability and robustness during restoration.

Let $p_\theta(y|x)$ denote the distribution over model prediction y, parameterized by $\theta \in \mathbb{R}^{|\theta|}$, given an input x. The significance of a parameter can be determined by measuring the impact its perturbation has on the model's output. The KL divergence $\mathcal{L}_{kl}(p_\theta(y|x) \parallel p_{\theta+\delta}(y|x))$ can be used to measure the sensitivity of this distribution to a small parameter perturbation $\delta \in \mathbb{R}^{|\theta|}$. Martens (2020), Pascanu and Bengio (2013) shows that as $\delta \to 0$, the following second-order approximation holds:

$$\mathbb{E}_x\left[\mathcal{L}_{kl}(p_\theta(y|x) \parallel p_{\theta+\delta}(y|x))\right] = \delta^T F_\theta \delta + O(\delta^3), \tag{9}$$

where $F_\theta \in \mathbb{R}^{|\theta|\times|\theta|}$ is the Fisher information matrix (FIM) (Fisher, 1922), defined as:

$$F_\theta = \mathbb{E}_x\left[\mathbb{E}_{y\sim p_\theta(y|x)} \nabla_\theta \log p_\theta(y|x) \nabla_\theta \log p_\theta(y|x)^T\right]. \tag{10}$$

This approximation demonstrates that the FIM links parameter perturbations δ to the resultant changes in the model's output distribution, indicating their importance. Consequently, we leverage FIM to guide the parameter restoration process. However, the dimensionality of FIM $|\theta| \times |\theta|$ makes it intractable to compute in practice. Therefore, consistent with prior work (Kirkpatrick et al., 2017), we adopt the diagonal approximation of FIM, represented as a vector in $\mathbb{R}^{|\theta|}$. Moreover, in practice, machine learning models are typically trained on a finite set of N training samples $\{(x_j, y_j)\}_{j=1}^N$, rather than having direct access to the true data distribution $p(x)$. In such case, the diagonal FIM can be empirically approximated:

$$\hat{F}_\theta = \frac{1}{N}\sum_{j=1}^{N} (\nabla_\theta \log p_\theta(y_j|x_j))^2, \tag{11}$$

where $\hat{F}_\theta \in \mathbb{R}^{|\theta|}$ is the empirical diagonal FIM. Each component of $\hat{F}_\theta$ corresponds to an individual parameter, and a larger value for such a

component signifies greater influence of the corresponding parameter on the model's predictions.

For CTTA, given a batch of N pairs of unlabeled test inputs and corresponding teacher's pseudo-label $\{(x_j^t, \hat{y}_j^t)\}_{j=1}^N$ at time step t, the FIM for the student model is defined as:

$$\hat{F}_{\theta_S^{t-1}} = \frac{1}{N}\sum_{j=1}^{N}(\nabla_{\theta_S^{t-1}}\log p_{\theta_S^{t-1}}(\hat{y}_j^t|x_j^t))^2. \tag{12}$$

Since the FIM is derived from pseudo-labels, which can be noisy in the target domain, the FIM itself might assign high importance to parameters associated with these erroneous predictions. Therefore, given a specific layer with parameters (still denoted as $\hat{\theta}_S^{t-1}$ for simplicity), ARR integrates a stochastic competent by employing a random matrix $R^t \sim \mathrm{Uniform}(0,1)$ with the same shape as the parameters, where each element follows a uniform distribution between 0 and 1. The reset scores W^t are then obtained by an element-wise multiplication of the FIM and the random matrix. The parameters are updated through:

$$W^t = \hat{F}_{\theta_S^{t-1}} \odot R^t, \tag{13}$$

$$M^t = W^t < \eta, \tag{14}$$

$$\theta_S^t = M^t \odot \theta^0 + (1 - M^t) \odot \hat{\theta}_S^{t-1}, \tag{15}$$

where $\hat{\theta}_S^{t-1}$ denotes the student model parameters after gradient descent from θ_S^{t-1} at time step t, $\odot$ is the element-wise multiplication, M^t indicates the mask matrix, $<$ represents element-wise less than operation, and η is the threshold value acquired by the q-quantile of W^t: $\eta = quantile(W^t, q)$. Consequently, the elements in M^t are set to 1, when the corresponding score value is less than η, indicating the associated parameters should be reset to the source parameters θ^0. This strategy enables the model to retain essential knowledge while introducing randomness to enhance robustness to mitigate forgetting.

4. Experimental setups

In this study, we rigorously evaluate our methodology across four benchmark tasks in CTTA object detection. These tasks encompass continual adaptation to synthetic and real-world distribution shifts, evaluated over short and long-term periods. Inspired by the foundational work CoTTA (Wang et al., 2022), the short-term CTTA tasks entails the sequential adaptation to various target domains once. In contrast, the long-term CTTA tasks involves continually adapting the model toward a group of target domains cyclically.

4.1. Datasets

Cityscapes. The Cityscapes (Cordts et al., 2016) is collected for urban scene understanding, encompassing 2975 training images and 500 validation images with eight categories, i.e. person, rider, car, truck, bus, train, motorcycle, and bicycle. We utilize the model pre-trained on this training set as the source model, and the data is discarded during adaptation.

Cityscapes-C. Hendrycks and Dietterich (2019) initially design to assess robustness against various corruptions, introducing 15 types of corruption with 5 severity levels. We create Cityscapes-C by applying these corruptions at the maximum severity level to the validation set of clean cityscapes, treating each corruption as an individual target domain comprising 500 images. Our short-term CTTA tasks selectively focuses on the latter 12 corruptions, including Defocus Blur, Frosted Glass Blur, Motion Blur, Zoom Blur, Snow, Frost, Fog, Brightness, Contrast, Elastic, Pixelate, and JPEG. For the long-term CTTA task, we prioritize the five corruptions related to autonomous driving scenarios as the target domain group following (Gan et al., 2023b), namely Fog, Motion, Snow, Brightness, and Defocus. We repeat adaptation to the target domain group 10 times to evaluate long-term performance.

SHIFT. The SHIFT (Sun et al., 2022) is a synthetic dataset for autonomous driving, featuring real-world environmental changes. SHIFT can be categorized as clear, cloudy, overcast, rainy, and foggy, where each condition contains images taken at various times ranging from daytime to night. For SHIFT, short-term CTTA tasks are considered. We designate the clear condition as the source domain, with the remaining four conditions as the target domain groups including nearly 20k images in total.

ACDC. The ACDC (Sakaridis et al., 2021) shares the same class types as Cityscapes and is collected in four different adverse visual conditions, including Fog, Night, Rain, and Snow. Following (Wang et al., 2022), we use these four conditions as the target domain group for the long-term CTTA task, with 400 unlabeled images per condition. Similarly, the source model is continually adapted to the target domain group for 10 cycles.

4.2. Implementation details

We adopt the Faster R-CNN with ResNet50 (He et al., 2016) pretrained on ImageNet (Krizhevsky et al., 2012) as the backbone. Following (VS et al., 2023b), we maintain a batch size of 1 to emulate a real-world application scenario where the detector adapts toward a continuous influx of images. The source models are trained using an SGD optimizer with a learning rate of 0.001 and a momentum of 0.9. Algorithms are implemented leveraging the Detectron2 (Wu et al., 2019). The metric of mAP at an IoU threshold of 0.5 (mAP0.5) is employed for evaluation. Each experiment is conducted on 1 NVIDIA A800 GPU.

4.3. Baselines and compared approaches

We compare AMROD with seven source-free baselines across various settings, including Source (Ren et al., 2015), Tent (Wang et al., 2020), IRG (VS et al., 2023a), MemCLR (VS et al., 2023b), CoTTA (Wang et al., 2022), SVDP (Yang et al., 2024), WHW (Yoo et al., 2024), and AMROD-upstop. Specifically, "Source" represents the source model without adaptation. Tent (Wang et al., 2020) updates the affine parameters through entropy minimization in TTA. Furthermore, Memclr and IRG are SFDA object detection methods, also referred to as TTA methods (Wang et al., 2022). MemCLR integrates cross-attention with CL, while IRG incorporates instance relation graph and CL. CoTTA utilizes a fixed threshold for pseudo-labeling and random neuron recovery to tackle CTTA, SVDP explores sparse visual prompts for CTTA dense prediction, and WHW utilizes source feature statistics to determine whether to update the lightweight adapter integrated into the backbone to address CTTA. "AMROD-unstop" represents AMROD without dynamic skipping.

5. Results and analysis

5.1. Synthetic continual distribution shift

5.1.1. Short-term CTTA tasks results

We first evaluate the effectiveness of AMROD on the short-term Cityscapes-to-Cityscapes-C adaptation tasks. As depicted in Table 2, Tent (Wang et al., 2020) undergoes a slight decline in performance, dropping from 15.1 to 14.2 mAP relative to the source model. This downturn may be attributed to its dependency on a large batch size to update the parameters of the batchnorm layer, making it suboptimal for online adaptation. In contrast, CoTTA, IRG, and Memclr exhibit enhancements in performance, achieving 16.0, 17.5, and 18.0 mAP respectively. Although the SFDA methods like IRG and MemCLR assume a static target domain, employing a large momentum update rate α for the teacher enables a relatively stable adaptation in the short term. Consequently, CL leads to a better performance than CoTTA which solely employs weight-averaged predictions. Moreover, SVDP, which utilizes

Table 2
Experimental results (mAP0.5) and adaptation iterations of Cityscapes-to-Cityscapes-C short-term CTTA task. We evaluate the performance by continually adapting the source model to twelve corruptions. "-unstop" denotes AMROD without skipping.

Time	t ⟶												All		
Condition	Defocus	Glass	Motion	Zoom	Snow	Frost	Fog	Brightness	Contrast	Elastic	Pixelate	Jpeg	Mean	Gain	Iter.
Source	6.8	8.1	8.0	1.5	0.2	6.8	34.6	30.7	3.0	50.2	17.6	13.5	15.1	/	/
Tent	6.8	7.8	7.7	1.3	0.2	6.1	33.1	28.0	2.2	51.1	14.8	11.0	14.2	–0.9	6.0k
CoTTA	7.8	9.0	8.9	1.8	0.3	7.1	38.4	31.1	8.6	49.6	16.2	13.1	16.0	+0.9	6.0k
SVDP	7.7	10.1	9.7	2.3	0.7	13.0	42.4	45.2	15.4	47.2	21.2	14.8	19.1	+4.0	6.0k
IRG	8.0	11.0	9.3	3.4	1.2	13.0	37.9	41.3	15.9	38.9	16.9	13.4	17.5	+2.4	6.0k
MemCLR	8.5	10.4	10.6	2.7	1.1	12.2	41.4	41.6	16.4	43.1	15.4	12.7	18.0	+2.9	6.0k
WHW	**9.2**	9.3	12.2	1.9	0.8	14.9	44.6	**48.9**	10.5	**51.9**	20.0	15.3	20.0	+4.9	4.1k
Ours-unstop	8.6	**12.1**	11.7	**3.6**	**1.5**	**16.7**	**44.7**	48.1	16.7	47.4	22.5	13.9	20.6	+5.5	6.0k
Ours	8.7	11.7	**12.4**	**3.6**	**1.5**	13.5	43.2	47.3	**18.4**	47.0	**24.8**	**17.2**	**20.8**	**+5.7**	**4.7k**

Table 3
Experimental results (mAP0.5) and adaptation iterations of Cityscapes-to-Cityscapes-C long-term CTTA task. We evaluate the performance by continually adapting the source model to the five corruption ten times. "-unstop" denotes AMROD without skipping. To save space, we display selected rounds of results.

Time	t ⟶																	
Round	1					5					10					All		
Condition	Fog	Motion	Snow	Brightness	Defocus	Fog	Motion	Snow	Brightness	Defocus	Fog	Motion	Snow	Brightness	Defocus	Mean	Gain	Iter.
Source	36.1	8.1	0.2	31.0	6.7	36.1	8.1	0.2	31.0	6.7	36.1	8.1	0.2	31.0	6.7	16.4	/	/
Tent	35.8	8.1	0.2	29.2	6.2	30.7	10.2	0.8	27.9	12.4	20.4	5.0	0.1	11.5	2.4	12.0	–4.4	25.0k
CoTTA	38.2	**10.6**	0.4	33.8	9.3	40.8	10.4	0.6	36.5	9.2	40.8	10.4	0.5	36.3	9.6	19.1	+2.7	25.0k
SVDP	36.8	8.8	0.6	43.5	11.2	45.0	15.4	3.8	47.8	19.9	41.5	16.2	4.6	44.6	20.2	25.3	+8.9	25.0k
IRG	37.9	9.0	0.7	44.3	12.2	45.6	17.7	6.0	46.7	21.9	37.0	16.7	7.1	38.5	20.5	25.6	+9.2	25.0k
MemCLR	37.7	8.9	**0.8**	45.5	13.5	45.0	18.1	5.1	46.8	22.4	36.8	18.5	7.5	38.4	21.8	26.0	+9.6	25.0k
WHW	**40.8**	9.9	0.7	43.4	12.5	42.8	15.5	2.0	46.9	17.9	36.7	18.3	3.5	44.1	16.2	24.2	+7.8	18.0k
Ours-unstop	39.0	10.4	**0.8**	**48.0**	**13.8**	**49.4**	**18.6**	7.7	51.7	**25.0**	45.7	**20.0**	**11.8**	46.4	25.8	29.0	+12.6	25.0k
Ours	39.2	9.9	**0.8**	46.7	12.7	**49.4**	18.5	**8.0**	**52.9**	24.7	**46.2**	19.5	11.3	**47.3**	**29.1**	**29.2**	**+12.8**	**20.1k**

Table 4
Experimental results (mAP0.5) of SHIFT short-term CTTA task. We evaluate the performance by continually adapting the source model to the four conditions.

Time	t ⟶				All		
Condition	Cloudy	Overcast	Rainy	Foggy	Mean	Gain	Iter.
Source	51.8	41.5	43.8	33.9	42.7	/	/
Tent	50.9	39.5	36.3	23.1	37.5	–5.2	19.6k
CoTTA	51.1	40.1	40.7	29.9	40.5	–2.2	19.6k
SVDP	52.0	41.0	43.8	35.9	43.2	+0.5	19.6k
IRG	51.9	40.6	42.7	34.3	42.4	–0.3	19.6k
MemCLR	51.8	40.4	42.6	35.0	42.4	–0.3	19.6k
WHW	51.7	41.4	43.9	34.0	42.8	+0.1	10.2k
Ours-unstop	**52.3**	41.3	44.1	36.8	43.6	+0.9	19.6k
Ours	**52.3**	**41.6**	**44.4**	**37.3**	**43.9**	**+1.2**	**16.4k**

visual prompts, achieves a sub-optimal performance of 19.1 mAP. Furthermore, while WHW selectively updates lightweight adapters using fewer adaptation iterations, this strategy might skip necessary adaptation steps, thereby limiting its performance to 20.0 mAP. Remarkably, our proposed method improves the performance to 20.8 mAP, consistently outperforming the above approaches. Additionally, the dynamic skipping strategy in AM reduces adaptation iterations by 20% while achieving a 0.2 mAP improvement compared to AMROD-unstop. AMROD ensures a more reliable and efficient adaptation to target domains characterized by intense changes in the short term.

5.1.2. Long-term CTTA tasks results

As presented in Table 3, the long-term tasks outcomes reveal the source model's poor performance, with an average mAP of 16.4. Despite the improvement of Memclr and IRG, their performance also begins to decline in the later rounds, failing to maintain stability over long-term adaptation. We believe this is due to the aforementioned methods not accounting for continual distribution shifts, resulting in error accumulation and catastrophic forgetting. Furthermore, CoTTA utilizes a stochastic restoration mechanism to mitigate forgetting, but its randomness may result in losing crucial information, thus limiting its performance at 19.1 mAP. We also observe that WHW exhibits noticeable performance degradation over prolonged continuous shifts, dropping to an average of 24.2 mAP. This indicates that freezing the backbone and relying solely on the additional adapters struggles to capture evolving target knowledge over extended periods. In contrast, SVDP which employs sparse visual prompts raises the performance to 25.3 mAP. Particularly, AMROD yields a remarkable 12.8 mAP enhancement over the Source and surpasses all comparative baselines using 80% adaptation iterations, which employs the ARR mechanism to conserve valuable knowledge while eliminating noise from prior domains. These findings empirically validate the effectiveness of our proposed method in ensuring stable and efficient adaptation amidst synthetic continual distribution shifts, across both short-term and long-term scenarios.

Table 5
Experimental results (mAP0.5) and adaptation iterations of Cityscapes-to-ACDC long-term CTTA task. We evaluate the performance by continually adapting the source model to the four conditions ten times. "-unstop" denotes AMROD without skipping. To save space, we display selected rounds of results.

Time	t ⟶																		
Round	1				4				7				10				All		
Condition	Fog	Night	Rain	Snow	Fog	Night	Rain	Snow	Fog	Night	Rain	Snow	Fog	Night	Rain	Snow	Mean	Gain	Iter.
Source	52.3	18.7	33.5	39.6	52.3	18.7	33.5	39.6	52.3	18.7	33.5	39.6	52.3	18.7	33.5	39.6	36.0	/	/
Tent	52.4	18.6	33.4	38.9	51.7	17.4	31.4	36.0	45.8	14.6	28.1	28.5	35.0	9.5	21.5	18.3	30.5	-5.5	16.0k
CoTTA	53.7	19.7	38.0	42.4	53.1	19.7	37.7	42.9	51.3	19.0	36.5	41.8	50.9	19.1	36.0	42.4	37.8	+1.8	16.0k
SVDP	52.8	20.0	35.6	42.0	54.6	23.5	38.7	43.8	52.8	24.0	38.6	43.9	51.8	23.6	38.2	43.0	39.5	+3.5	16.0k
IRG	52.7	20.6	36.0	42.9	53.6	23.2	38.1	45.6	51.0	23.1	37.1	42.8	49.7	22.5	36.5	40.7	38.9	+2.9	16.0k
MemCLR	52.9	**21.2**	35.2	42.8	52.8	23.2	38.2	44.4	51.4	23.2	37.1	42.9	49.9	22.9	36.6	41.0	38.7	+2.7	16.0k
WHW	**54.5**	21.1	**39.6**	**43.5**	**55.7**	21.3	39.1	41.5	53.0	21.7	35.9	39.5	50.9	20.9	33.8	38.3	38.3	+2.3	8.0k
Ours-unstop	52.9	20.4	34.5	42.9	54.5	23.8	**39.6**	**45.8**	**53.5**	25.1	**39.0**	45.3	52.5	24.3	**38.8**	44.7	40.3	+4.3	16.0k
Ours	52.8	20.1	35.2	42.9	54.4	**24.6**	38.6	45.5	**53.5**	**25.6**	38.5	**46.2**	**53.5**	**25.4**	37.4	**45.3**	**40.4**	**+4.4**	**13.2k**

5.2. Real-world continual distribution shift

5.2.1. Short-term CTTA tasks results

We also evaluate our method on the real-world continual distribution shift dataset. The results of the SHIFT short-term CTTA tasks are shown in Table 4. While other baseline methods suffer from performance decline, SVDP which employs a fixed threshold for pseudo-labeling, stochastic restoration, and prompts learning achieves a 0.5 mAP improvement. Additionally, while WHW uses fewer adaptation iterations, it risks skipping necessary adaptations, resulting in a marginal improvement of 0.1 mAP. Moreover, AMROD which introduces the AM module shows a superior performance of 43.9 mAP with 16% improvement in efficiency, achieving a better parameter-efficiency tradeoff.

5.2.2. Long-term CTTA tasks results

The experimental results of the ACDC long-term CTTA tasks are summarized in Table 5. CoTTA, IRG, and MemCLR potentially suffer from error accumulation and catastrophic forgetting, manifesting in a rapid decline in performance during the later stages. Additionally, WHW also struggles with long-term shifts, as its performance under Fog conditions drops from 54.5 mAP in the beginning to 50.9 mAP at the end. Similarly, SVDP achieves a sub-optimal performance of 39.5 mAP. Notably, AMROD achieves a performance of 40.4 mAP, yielding an absolute improvement of 0.9 mAP over baselines. While AMROD may not display a significant advantage over other methods in the initial round, its performance continually improves and remains stable throughout the long-term adaptation.

These results underscore the capability of our method to foster robust adaptation toward the real-world continual distribution shift in both the short and long term.

5.3. Qualitative results

As shown in Figs. 4 and 5, we compare the source model, baseline methods, and AMROD in the final adaptation to the brightness and defocus on the long-term Cityscapes-to-Cityscapes-C tasks. AMROD guides the model in learning better feature representations while effectively mitigating forgetting in the long-term adaptation. Therefore, AMROD assists detector in distinguishing more foreground object categories and better locating them.

5.4. Ablation study

5.4.1. Ablation study results on model components

We conduct an ablation study to evaluate the impact of the OCL, AM, and ARR. As shown in Table 6, integrating the mean-teacher architecture achieves the performance of 25.9 mAP. Building upon this foundation, the proposed OCL, AM, and ARR achieve an additional 1.5, 1.4, and 0.8 mAP improvement individually. Moreover, the combination of the OCL and AM modules raises the performance to 28.6 mAP. Furthermore, including ARR mechanism brings a 0.6 mAP boost in performance, culminating in a model performance of 29.2 mAP.

Table 6
Ablation experiment on Model Components. "Mean-Teacher" represents the base mean teacher framework with weak-strong augmentation and KL divergence distillation. All experiments are done on long-term Cityscapes-to-Cityscapes-C tasks.

Mean-Teacher	OCL	AM	ARR	Mean	Gain
✓				25.9	/
✓			✓	26.7	+0.8
✓		✓		27.3	+1.4
✓	✓			27.4	+1.5
✓	✓	✓		28.6	+2.7
✓	✓	✓	✓	29.2	+3.3

Table 7
Ablation Study on Module Variants. The "student" variant utilizes predictions from the student model. The "FT" represents replacing AM module in AMROD with a fixed threshold. The "SR" and the "DR" indicate replacing ARR with the stochastic restoration (Wang et al., 2022) and the data-driven restoration (Brahma & Rai, 2023), respectively. All experiments are done on long-term Cityscapes-to-Cityscapes-C tasks.

AMROD	Student	FT of 0.9	FT of 0.8	FT of 0.7	SR	DR	Mean
✓	✓						27.9
✓		✓					27.7
✓			✓				28.2
✓				✓			28.5
✓					✓		28.7
✓						✓	28.3
✓							**29.2**

5.4.2. Ablation study results on module variants

As presented in Table 7, we further perform a fine-grained comparison study to assess the effectiveness of the modules of AMROD. This is done by replacing the teacher predictions with the student prediction in the Mean-Teacher framework (Tarvainen & Valpola, 2017), AM with a Fixed threshold (FT) strategy, and ARR with stochastic restoration (SR) (Wang et al., 2022) or data-driven restoration (DR) (Brahma & Rai, 2023) mechanism. The findings indicate a variant utilizing student model predictions achieved only 27.9 mAP. This performance dip is likely because the student model, being actively updated, is less stable than the teacher model, which is updated via the EMA strategy to mitigate error accumulation. Furthermore, replacing the AM module with FT strategies (FT of 0.9, 0.8, and 0.7) resulted in mAP scores of

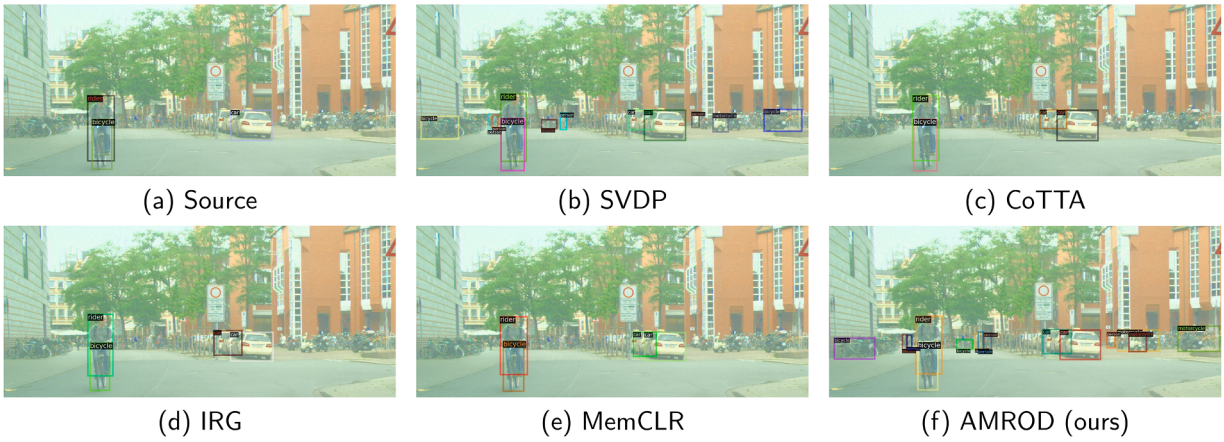

Fig. 4. Qualitative results. We compare the detection results of the AMROD and other baseline methods in the 10th round of adaption to *Brightness* corruption on the long-term Cityscapes-to-Cityscapes-C task.

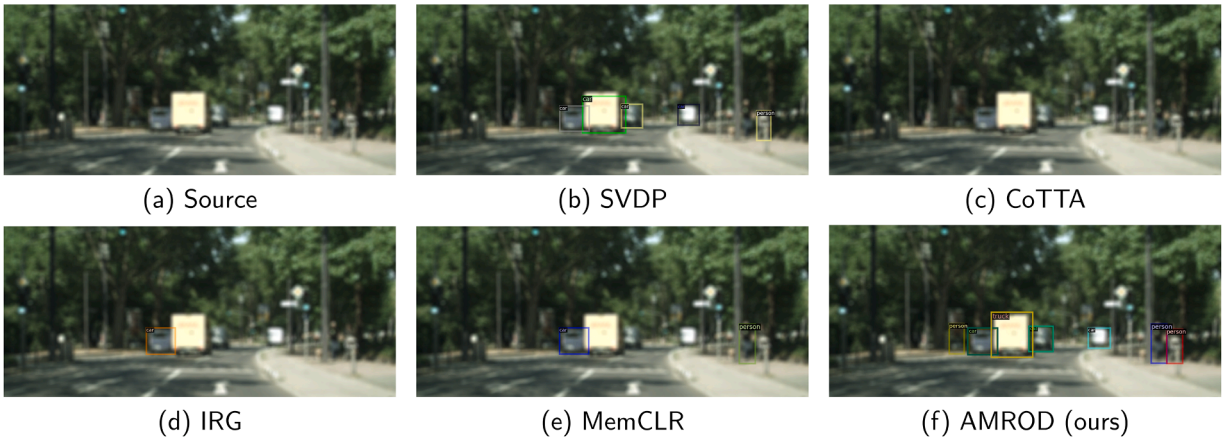

Fig. 5. Qualitative results. We compare the detection results of the AMROD and other baseline methods in the 10th round of adaption to *Defocus* corruption on the long-term Cityscapes-to-Cityscapes-C task.

27.7, 28.2, and 28.5, respectively. These are notably lower than AMROD's 29.2 mAP, underscoring the benefit of AM's adaptability to varying model confidence across diverse categories and domains, which FT inherently lack. Similarly, when the ARR module is substituted with SR or DR, performance drop to 28.7 mAP and 28.3 mAP, respectively. This suggests that SR's randomness might discard valuable domain-specific knowledge, while DR could retain noisy parameters. In contrast, AMROD's ARR, which selectively resets inactive parameters with higher probability while integrating a stochastic component, better preserves essential knowledge and ensures robust adaptation.

5.4.3. Ablation study results on hyperparameter impact

To investigate the impact of hyperparameters, we conduct an ablation study on six key hyperparameters on the long-term Cityscapes-to-Cityscapes-C task, revealing the model's sensitivity and optimal configurations over the five target domains. These hyperparameters are varied while keeping others at their default settings, including β_t, δ^0, ϵ, β_s, δ_s, and η. As shown in Fig. 6, for the AM module, the dynamic threshold update rate β_t (Fig. 6a) shows peak performance at 0.95, suggesting that a balanced reliance on recent and historical category scores is beneficial. The initial threshold δ^0 (Fig. 6b) demonstrates strong robustness, while the linear projection factor ϵ (Fig. 6c) is optimal at 1.3, indicating that a moderate scaling of predicted scores best guides threshold adaptation. For dynamic skipping, the update rate β_s (Fig. 6d) achieves its best at 0.75 and the stability range parameter β_s (Fig. 6e) performs best at 1.3, effectively balancing the reaction to mean score fluctuations for pausing adaptation. Finally, the threshold η (Fig. 6f) in ARR, which determines the q-quantile for resetting parameters, yields peak performance at 0.0001. Deviations from this value likely result in either insufficient resetting of inactive parameters or excessive resetting of potentially useful knowledge. These findings highlight that the selected hyperparameters offer a balanced performance and confirm the stability of AMROD, as they demonstrate minimal sensitivity to small hyperparameter variations.

5.5. Discussion

Firstly, following prior works on TTA in objection detection (VS et al., 2023a,b), our experiments are primarily based on a single detector backbone, i.e., Faster R-CNN (Ren et al., 2015). Additionally, AMROD is specifically designed for object detection tasks and has demonstrated its effectiveness exclusively in this domain. Future work will extend to other advanced detector backbone networks or other vision tasks to assess the universality and generalization capabilities of AMROD. Moreover, the proposed ARR mechanism to reset the model every iteration increases the computational cost. Future work could explore more efficient mechanisms for restoration, such as restoring the model only when significant changes in the distribution of the target domain are detected.

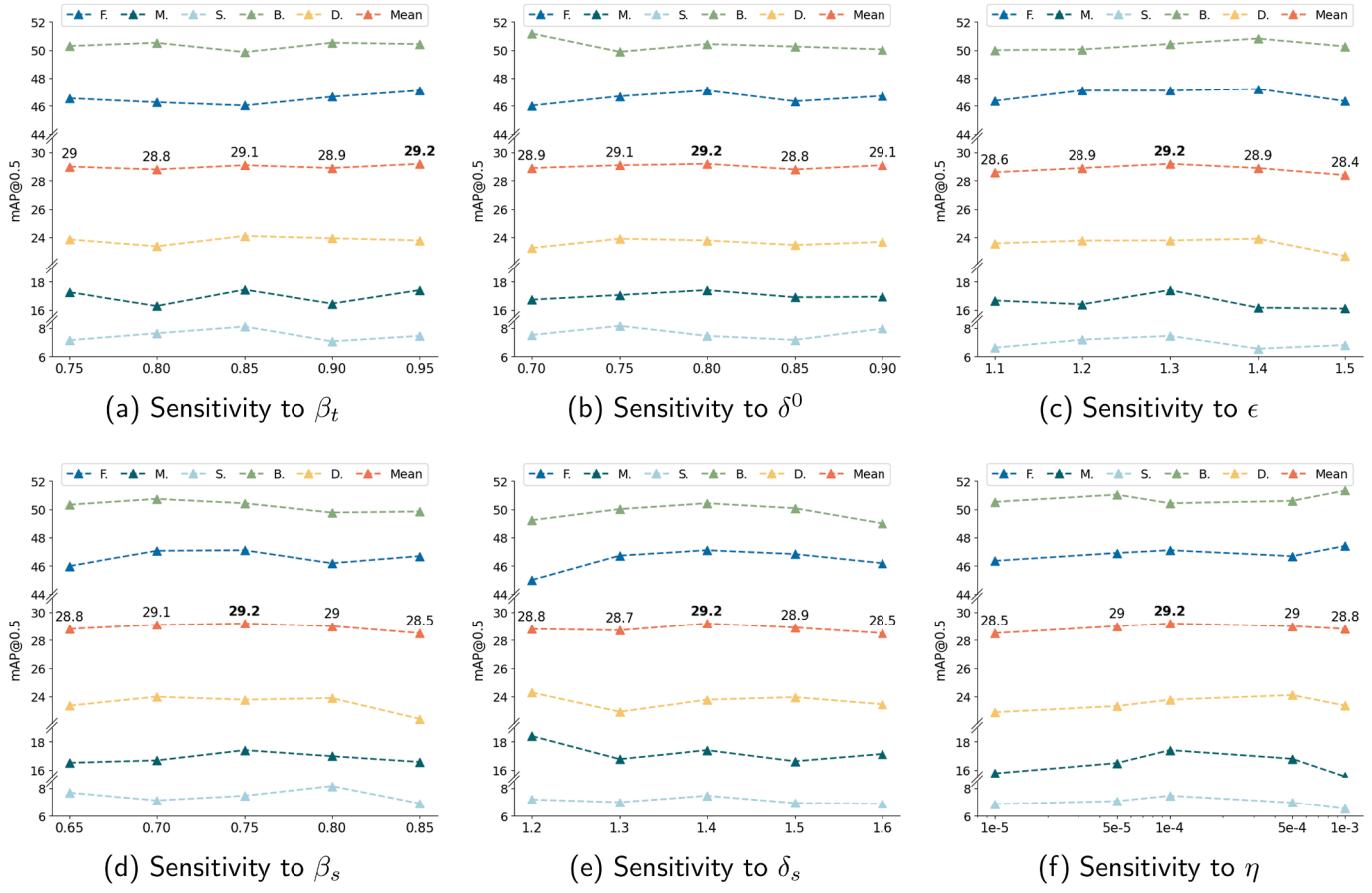

Fig. 6. Ablation Study on Hyperparameter Impact. The hyperparameters are: β_t, δ^0, and ϵ for dynamic threshold; β_s and δ_s for dynamic skipping in AM; and η for ARR. The "F.", "M.", "S.", "B.", and "D." represent the average performance over 10 times adaptation of Fog, Motion, Snow, Brightness, and Defocus, respectively. All experiments are done on long-term Cityscapes-to-Cityscapes-C tasks. .

Finally, the evaluation tasks in our work are designed to simulate real-world adaptation scenarios by incorporating datasets affected by corruption and adverse conditions. However, real-world data distributions are inherently more complex. Consequently, a promising avenue for future research would be to apply our methodology in practical, real-world systems to further validate its effectiveness.

6. Conclusion

In this work, we propose AMROD to address the two challenges in Continual Test-Time Adaptation (CTTA). Firstly, object-level contrastive learning leverages ROI features for contrastive learning to refine the feature representation, tailored for object detection. Secondly, the adaptive monitoring module enables efficient adaptation and high-quality pseudo-labels by dynamically skipping unstable adaptation and updating the category-specific threshold, based on the predicted confidence scores. Lastly, the adaptive randomized restoration mechanism selectively resets inactive parameters to mitigate forgetting while retaining valuable knowledge. Extensive empirical evaluations across four CTTA benchmarks demonstrate the effectiveness of AMROD for both short-term and long-term adaptation under synthetic and real-world continual distribution shifts. Notably, AMROD achieves state-of-the-art results on the long-term Cityscapes-to-Cityscapes-C task, improving accuracy by up to 3.2 mAP over existing baselines while simultaneously enhancing computational efficiency. Ultimately, this work provides a robust, source-free framework capable of maintaining high-performance object detection in complex, non-stationary environments encountered by real-world perception systems.

CRediT authorship contribution statement

Shilei Cao: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Project administration, Writing – original draft, Writing – review & editing, Visualization; **Juepeng Zheng:** Supervision, Writing – original draft, Writing – review & editing, Funding acquisition, Resources, Supervision; **Yan Liu:** Data curation, Formal analysis, Software, Visualization, Writing – original draft, Writing – review & editing; **Baoquan Zhao:** Writing – review & editing; **Ziqi Yuan:** Writing – review & editing; **Weijia Li:** Writing – review & editing; **Runmin Dong:** Writing – review & editing; **Haohuan Fu:** Funding acquisition, Resources, Supervision.

Data availability

Data will be made available on request.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

This work was supported by the Guangdong Science and Technology Program (2024B0101040005), National Natural Science Foundation of China (Grant No. T2125006 and No. 42401415), Shenzhen Science and Technology Program (KCXFZ20240903093759004 and

KJZD20230923115106012), and Guangdong Science and Technology Program (2025B0101080001) (Corresponding author: Juepeng Zheng).

References

Aljundi, R., Chakravarty, P., & Tuytelaars, T. (2017). Expert gate: Lifelong learning with a network of experts. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 3366–3375).

An, S.-Y., Kang, J.-G., Choi, W.-S., & Oh, S.-Y. (2011). A neural network based retrainable framework for robust object recognition with application to mobile robotics. *Applied Intelligence*, *35*(2), 190–210.

Brahma, D., & Rai, P. (2023). A probabilistic framework for lifelong test-time adaptation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 3582–3591).

Buciluǎ, C., Caruana, R., & Niculescu-Mizil, A. (2006). Model compression. In *Proceedings of the 12th ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 535–541).

Cao, S., Gong, Z., Lin, H., Liu, Y., Cheng, J., Hu, X., Liang, H., Li, G., Qin, C., Cheng, H. et al. (2025). Crossearth-gate: Fisher-guided adaptive tuning engine for efficient adaptation of cross-domain remote sensing semantic segmentation. arXiv preprint arXiv:2511.20302.

Chen, G., Choi, W., Yu, X., Han, T., & Chandraker, M. (2017). Learning efficient object detection models with knowledge distillation. *Advances in Neural Information Processing Systems*, *30*.

Chen, T., Kornblith, S., Norouzi, M., & Hinton, G. (2020). A simple framework for contrastive learning of visual representations. In *International conference on machine learning* (pp. 1597–1607). PMLR.

Chen, Z., Wang, Z., & Zhang, Y. (2023). Exploiting low-confidence pseudo-labels for source-free object detection. In *Proceedings of the 31st ACM international conference on multimedia* (pp. 5370–5379).

Cordts, M., Omran, M., Ramos, S., Rehfeld, T., Enzweiler, M., Benenson, R., Franke, U., Roth, S., & Schiele, B. (2016). The cityscapes dataset for semantic urban scene understanding. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 3213–3223).

De Lange, M., Aljundi, R., Masana, M., Parisot, S., Jia, X., Leonardis, A., Slabaugh, G., & Tuytelaars, T. (2021). A continual learning survey: Defying forgetting in classification tasks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *44*(7), 3366–3385.

Döbler, M., Marsden, R. A., & Yang, B. (2023). Robust mean teacher for continual and gradual test-time adaptation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 7704–7714).

Doulamis, A. D., Doulamis, N. D., & Kollias, S. D. (2000). On-line retrainable neural networks: Improving the performance of neural networks in image analysis problems. *IEEE Transactions on Neural Networks*, *11*(1), 137–155.

Doulamis, N., Doulamis, A., & Ntalianis, K. (2002). Recursive non-linear autoregressive models (RNAR): Application to traffic prediction of MPEG video sources. In *2002 11th European signal processing conference* (pp. 1–4). IEEE.

Fisher, R. A. (1922). On the mathematical foundations of theoretical statistics. *Philosophical transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, *222*(594–604), 309–368.

Fu, R., Han, J., Sun, Y., Wang, S., Al-Absi, M. A., Wang, X., & Sun, H. (2025). Robust crop disease detection using multi-domain data augmentation and isolated test-time adaptation. *Expert Systems with Applications*, (127324).

Gan, Y., Bai, Y., Lou, Y., Ma, X., Zhang, R., Shi, N., & Luo, L. (2023a). Decorate the newcomers: Visual domain prompt for continual test time adaptation. In *Proceedings of the AAAI conference on artificial intelligence* (pp. 7595–7603). (*vol. 37*).

Gan, Y., Pan, M., Zhang, R., Ling, Z., Zhao, L., Liu, J., & Zhang, S. (2023b). Cloud-device collaborative adaptation to continual changing environments in the real-world. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 12157–12166).

Gong, T., Jeong, J., Kim, T., Kim, Y., Shin, J., & Lee, S.-J. (2022). Note: Robust continual test-time adaptation against temporal correlation. *Advances in Neural Information Processing Systems*, *35*, 27253–27266.

He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). Mask R-CNN. In *Proceedings of the IEEE international conference on computer vision* (pp. 2961–2969).

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770–778).

Hendrycks, D., & Dietterich, T. (2019). Benchmarking neural network robustness to common corruptions and perturbations. arXiv preprint arXiv:1903.12261.

Hinton, G., Vinyals, O., & Dean, J. (2015). Distilling the knowledge in a neural network. arXiv preprint arXiv:1503.02531.

Hoffman, J., Darrell, T., & Saenko, K. (2014). Continuous manifold based adaptation for evolving visual domains. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 867–874).

Hu, X., Uzunbas, G., Chen, S., Wang, R., Shah, A., Nevatia, R., & Lim, S.-N. (2021). Mixnorm: Test-time adaptation through online normalization estimation. arXiv preprint arXiv:2110.11478.

Huang, J., Guan, D., Xiao, A., & Lu, S. (2021). Model adaptation: Historical contrastive learning for unsupervised domain adaptation without source data. *Advances in Neural Information Processing Systems*, *34*, 3635–3649.

Ioannou, S., Kessous, L., Caridakis, G., Karpouzis, K., Aharonson, V., & Kollias, S. (2006). Adaptive on-line neural network retraining for real life multimodal emotion recognition. In *International conference on artificial neural networks* (pp. 81–92). Springer.

Iwasawa, Y., & Matsuo, Y. (2021). Test-time classifier adjustment module for model-agnostic domain generalization. *Advances in Neural Information Processing Systems*, *34*, 2427–2440.

Jiang, J., Zhao, S., Zhu, J., Tang, W., Xu, Z., Yang, J., Liu, G., Xing, T., Xu, P., & Yao, H. (2025). Multi-source domain adaptation for panoramic semantic segmentation. *Information Fusion*, *117*, 102909.

Kirkpatrick, J., Pascanu, R., Rabinowitz, N., Veness, J., Desjardins, G., Rusu, A. A., Milan, K., Quan, J., Ramalho, T., Grabska-Barwinska, A. et al. (2017). Overcoming catastrophic forgetting in neural networks. *Proceedings of International Conference on Computer Vision*, *114*(13), 3521–3526.

Kollias, D., Tagaris, A., & Stafylopatis, A. (2016). On line emotion detection using retrainable deep neural networks. In *2016 IEEE Symposium series on computational intelligence (SSCI)* (pp. 1–8). IEEE.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*, *25*, 248–255.

Kullback, S., & Leibler, R. A. (1951). On information and sufficiency. *The Annals of Mathematical Statistics*, *22*(1), 79–86.

Li, S., Ye, M., Zhu, X., Zhou, L., & Xiong, L. (2022a). Source-free object detection by learning to overlook domain style. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 8014–8023).

Li, Y.-J., Dai, X., Ma, C.-Y., Liu, Y.-C., Chen, K., Wu, B., He, Z., Kitani, K., & Vajda, P. (2022b). Cross-domain adaptive teacher for object detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 7581–7590).

Li, Z., Geng, L., Liu, Y., Rong, F., Ma, M., Tong, J., & Xiao, Z. (2025). Uncertainty-guided denoising bi-classifier adversarial domain adaptation network for cross-domain fault diagnosis. *Expert Systems with Applications*, (129742).

Li, Z., & Hoiem, D. (2017). Learning without forgetting. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *40*(12), 2935–2947.

Liang, H., Cao, S., Lai, Y., & Zheng, J. (2025a). Federated open-set domain generalization with adaptive adjustment boundary and weights. In *2025 IEEE International conference on multimedia and expo (ICME)* (pp. 1–6). IEEE.

Liang, H., Zhang, X., Cao, S., Li, G., & Zheng, J. (2025b). TTA-FEDDG: Leveraging test-time adaptation to address federated domain generalization. In *Proceedings of the AAAI conference on artificial intelligence* (pp. 18658–18666). (*vol. 39*).

Liang, Y., Cao, S., Zheng, J., Zhang, X., Huang, J., & Fu, H. (2025c). Low saturation confidence distribution-based test-time adaptation for cross-domain remote sensing image classification. *International Journal of Applied Earth Observation and Geoinformation*, *139*, 104463.

Liu, J., Yang, S., Jia, P., Zhang, R., Lu, M., Guo, Y., Xue, W., & Zhang, S. (2023). Vida: Homeostatic visual domain adapter for continual test time adaptation. arXiv preprint arXiv:2306.04344.

Liu, Y.-C., Ma, C.-Y., He, Z., Kuo, C.-W., Chen, K., Zhang, P., Wu, B., Kira, Z., & Vajda, P. (2021). Unbiased teacher for semi-supervised object detection. arXiv preprint arXiv:2102.09480.

Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 3431–3440).

Lü, S., Li, Z., Zhang, X., & Li, J. (2024). Consistency regularization-based mutual alignment for source-free domain adaptation. *Expert Systems with Applications*, *241*, 122577.

Martens, J. (2020). New insights and perspectives on the natural gradient method. *Journal of Machine Learning Research*, *21*(146), 1–76.

Ni, J., Yang, S., Liu, J., Li, X., Jiao, W., Xu, R., Chen, Z., Liu, Y., & Zhang, S. (2023). Distribution-aware continual test time adaptation for semantic segmentation. arXiv preprint arXiv:2309.13604.

Niloy, F. F., Ahmed, S. M., Raychaudhuri, D. S., Oymak, S., & Roy-Chowdhury, A. K. (2024). Effective restoration of source knowledge in continual test time adaptation. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision* (pp. 2091–2100).

Niu, S., Wu, J., Zhang, Y., Chen, Y., Zheng, S., Zhao, P., & Tan, M. (2022). Efficient test-time model adaptation without forgetting. In *International conference on machine learning* (pp. 16888–16905). PMLR.

Ntalianis, K., Doulamis, A., Doulamis, N., & Kollias, S. (2002). Unsupervised stereoscopic video object segmentation based on active contours and retrainable neural networks. *Signal Processing, Computational Geometry and Vision, World Scientific and Engineering Academy and Society Press*, *1*(1), 287–293.

Pascanu, R., & Bengio, Y. (2013). Revisiting natural gradient for deep networks. arXiv preprint arXiv:1301.3584.

Rebuffi, S.-A., Kolesnikov, A., Sperl, G., & Lampert, C. H. (2017). ICARL: Incremental classifier and representation learning. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2001–2010).

Reddy, N., Baktashmotlagh, M., & Arora, C. (2024). Towards domain-aware knowledge distillation for continual model generalization. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision* (pp. 696–707).

Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster R-CNN: Towards real-time object detection with region proposal networks. *Advances in Neural Information Processing Systems*, *28*. https://doi.org/10.48550/arXiv.1506.01497

Saito, K., Ushiku, Y., Harada, T., & Saenko, K. (2019). Strong-weak distribution alignment for adaptive object detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 6956–6965).

Saito, K., Watanabe, K., Ushiku, Y., & Harada, T. (2018). Maximum classifier discrepancy for unsupervised domain adaptation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 3723–3732).

Sakaridis, C., Dai, D., & Van Gool, L. (2021). ACDC: The adverse conditions dataset with correspondences for semantic driving scene understanding. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 10765–10775).

Song, J., Lee, J., Kweon, I. S., & Choi, S. (2023). Ecotta: Memory-efficient continual test-time adaptation via self-distilled regularization. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 11920–11929).

Sun, T., Segu, M., Postels, J., Wang, Y., Van Gool, L., Schiele, B., Tombari, F., & Yu, F. (2022). Shift: A synthetic driving dataset for continuous multi-task domain adaptation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 21371–21382).

Sun, Y., Wang, X., Liu, Z., Miller, J., Efros, A., & Hardt, M. (2020). Test-time training with self-supervision for generalization under distribution shifts. In *International conference on machine learning* (pp. 9229–9248). PMLR.

Szatkowski, F., Pyla, M., Przewięźlikowski, M., Cygert, S., Twardowski, B., & Trzciński, T. (2024). Adapt your teacher: Improving knowledge distillation for exemplar-free continual learning. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision* (pp. 1977–1987).

Tarvainen, A., & Valpola, H. (2017). Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. *Advances in Neural Information Processing Systems*, *30*. https://doi.org/10.48550/arXiv.1703.01780

Tian, K., Zhang, C., Wang, Y., Xiang, S., & Pan, C. (2021). Knowledge mining and transferring for domain adaptive object detection. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 9133–9142).

Tiwari, R., Killamsetty, K., Iyer, R., & Shenoy, P. (2022). GCR: Gradient coreset based replay buffer selection for continual learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 99–108).

VS, V., Oza, P., & Patel, V. M. (2023a). Instance relation graph guided source-free domain adaptive object detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 3520–3530).

VS, V., Oza, P., & Patel, V. M. (2023b). Towards online domain adaptive object detection. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision* (pp. 478–488).

Wang, D., Shelhamer, E., Liu, S., Olshausen, B., & Darrell, T. (2020). Tent: Fully test-time adaptation by entropy minimization. arXiv preprint arXiv:2006.10726.

Wang, Q., Fink, O., Van Gool, L., & Dai, D. (2022). Continual test-time domain adaptation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 7201–7211).

Wang, S., Liu, Y., Liu, Z., Yuan, X., Ji, Y., & Liang, P. (2025). Dit-SFDA: A source-free domain adaptation method for intelligent diagnosis of cardiovascular diseases with limited heart sound samples. *Expert Systems with Applications*, (128118).

Wang, Y., Hong, J., Cheraghian, A., Rahman, S., Ahmedt-Aristizabal, D., Petersson, L., & Harandi, M. (2024). Continual test-time domain adaptation via dynamic sample selection. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision* (pp. 1701–1710).

Wu, Y., Kirillov, A., Massa, F., Lo, W.-Y., & Girshick, R. (2019). Detectron2. https://github.com/facebookresearch/detectron2.

Xu, C., Song, Y., Zheng, Q., Wang, Q., & Heng, P.-A. (2025). Unsupervised multi-source domain adaptation via contrastive learning for EEG classification. *Expert Systems with Applications*, *261*, 125452.

Xu, J., & Zhu, Z. (2018). Reinforced continual learning. *Advances in Neural Information Processing Systems*, *31*.

Yang, S., Wu, J., Liu, J., Li, X., Zhang, Q., Pan, M., Gan, Y., Chen, Z., & Zhang, S. (2024). Exploring sparse visual prompt for domain adaptive dense prediction. In *Proceedings of the AAAI conference on artificial intelligence* (pp. 16334–16342). (*vol. 38*).

Ye, Z., Li, G., Liang, H., Wang, Z., Cao, S., Lai, Y., & Zheng, J. (2025). Quantifying samples with invariance for source-free class incremental domain adaptation. In *Proceedings of the 33rd ACM international conference on multimedia* (pp. 3340–3349).

Yoo, J., Lee, D., Chung, I., Kim, D., & Kwak, N. (2024). What how and when should object detectors update in continually changing test domains? In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 23354–23363).

You, F., Li, J., & Zhao, Z. (2021). Test-time batch statistics calibration for covariate shift. arXiv preprint arXiv:2110.04065.

Yu, Z., Li, J., Du, Z., Li, F., Zhu, L., & Yang, Y. (2023). Noise-robust continual test-time domain adaptation. In *Proceedings of the 31st ACM international conference on multimedia* (pp. 2654–2662).

Zeng, R., Deng, Q., Xu, H., Niu, S., & Chen, J. (2023). Exploring motion cues for video test-time adaptation. In *Proceedings of the 31st ACM international conference on multimedia* (pp. 1840–1850).

Zhang, D., Ye, M., Liu, Y., Xiong, L., & Zhou, L. (2022). Multi-source unsupervised domain adaptation for object detection. *Information Fusion*, *78*, 138–148.

Zhu, J., Bolsterlee, B., Chow, B. V. Y., Song, Y., & Meijering, E. (2023). Uncertainty and shape-aware continual test-time adaptation for cross-domain segmentation of medical images. In *International conference on medical image computing and computer-assisted intervention* (pp. 659–669). Springer.