An Adaptive Training Method for Continual Semantic Segmentation with Task Boundary Detection

1st Izwan bin Sahid
Universiti Malaysia Sabah
Kota Kinabalu, Malaysia
elfwoods1123@gmail.com

2nd Jiamin Li
Axis Design Studio Pte Ltd
Singapore
jiamin.li0501@hotmail.com

***Abstract*—Background and Motivation: Continual Semantic Segmentation (CSS) aims to enable models to sequentially learn new knowledge in dynamic environments, akin to human learning, without forgetting previously acquired capabilities. However, existing methods commonly face the dual challenges of "catastrophic forgetting" and "semantic drift." The root cause lies in the model's inability to distinguish the boundaries between new and old tasks, preventing adaptive adjustments in learning strategies and resulting in a trade-off between stability and plasticity.Method: This paper proposes an Adaptive Training method for Continual Semantic Segmentation with Task Boundary Detection (AT-TBD). The core of this method is the introduction of a Task Boundary Detection (TBD) module, which dynamically determines if a new task has been introduced into the data stream by monitoring changes in the entropy of the model's prediction outputs.Implementation: Based on the judgments from the TBD module, the model adaptively switches its training strategy between a "stable learning phase" and a "boundary adaptation phase." During the stable learning phase, the model focuses on optimizing performance for the current task. When a task boundary is detected, it switches to the boundary adaptation phase, prioritizing the consolidation of old knowledge through techniques such as parameter freezing and enhanced knowledge distillation, thereby achieving intelligent control over the learning pace.Core Conclusion: This paper presents the AT-TBD framework and provides fully specified implementation details and evaluation protocols for reproducible continual semantic segmentation. Full benchmark experiments on PASCAL VOC 2012 and ADE20K under class-incremental settings are currently in progress and will be reported in the revised version.Significance and Value: It suggests that explicitly perceiving task changes and adaptively adjusting learning strategies may better manage the model's knowledge update process, offering a promising direction for resolving the stability-plasticity dilemma. This method not only enhances the performance and robustness of CSS models but also provides theoretical and practical insights for building more intelligent lifelong learning systems that more closely mimic human learning mechanisms.**

***Keywords*—Continual Semantic Segmentation, Task Bound ary Detection, Adaptive Training, Catastrophic Forgetting, Se mantic Drift**

I. INTRODUCTION

Semantic segmentation, a fundamental task in computer vision, aims to assign a semantic class label to each pixel in an image. It plays a crucial role in key application areas such as autonomous driving, precision medicine, and remote sensing image analysis. Traditional semantic segmentation models are typically trained once on static, closed-set datasets, assuming that all classes are known during the training phase. However, in real-world applications, environments are dynamic, and models must possess the ability for continual or incremental learning, i.e., to continuously learn from new data streams without accessing the full history of data, thereby acquiring knowledge of new classes or domains [1][2][3].

While continual learning offers the potential for models to adapt to an open world, it also introduces a core challenge—catastrophic forgetting. When a neural network model learns a new task, its internal parameters are adjusted to fit the new data distribution, which often leads to the severe degradation of knowledge acquired from previous tasks. This problem is particularly pronounced in Continual Semantic Segmentation (CSS). Due to the pixel-level prediction nature of the task, models must not only prevent forgetting of old classes but also contend with the challenge of semantic drift. This refers to the dynamic changes in the definition of the background class across different learning stages, where foreground classes from previous steps may be considered background in later steps, and future new classes are also treated as background in the current step. This ambiguity can cause the model to confuse new and old classes with the background, further exacerbating performance decline.

To address these challenges, researchers have proposed various CSS methods, which can be broadly categorized into three types: replay-based methods that consolidate old knowledge by storing a small subset of old data or their generated variants; regularization-based methods that constrain the updates of important parameters by adding regularization terms to the loss function to protect old knowledge; and dynamic-architecture methods that allocate independent model parameters for new tasks to reduce knowledge conflicts.

However, most existing methods employ a fixed, one-size-fits-all learning strategy throughout the entire continual learning process. They apply the same intensity of knowledge retention constraints, regardless of whether the model is in a stable phase of learning a specific task or at the critical transition point between old and new tasks. This uniform approach overlooks a key fact: the required balance between plasticity (the ability to learn new knowledge) and stability (the ability to retain old knowledge) differs at

Corresponding Author: Jiamin Li, 8 EAN KIAM PLACE, Singapore, 429107, jiamin.li0501@hotmail.com

various learning stages. Particularly at the boundary moments of task transitions, the model is most susceptible to knowledge confusion and forgetting. Existing methods, lacking explicit perception of these task boundaries, are unable to implement more targeted protective measures during these critical periods, leading to insufficient knowledge consolidation. Conversely, during the stable learning phase within a task, overly strong regularization can inhibit the model's ability to fully learn the current task, resulting in a waste of resources. Therefore, the ambiguity of task boundaries is a key bottleneck limiting the performance of current CSS methods.

Based on this analysis, this paper proposes an Adaptive Training method for Continual Semantic Segmentation with Task Boundary Detection (AT-TBD). We argue that a more intelligent continual learning system should be able to autonomously perceive changes in the learning pace, i.e., to recognize the arrival of a new task. To this end, we have designed a Task Boundary Detection (TBD) module that identifies the emergence of new tasks by monitoring changes in the model's prediction uncertainty. Guided by the signals from the TBD module, the model can adaptively switch between two training strategies: during the stable learning phase within a task, it employs a conventional training strategy to optimize for the current task; at the boundary adaptation phase of a task transition, it activates stronger knowledge protection mechanisms, such as freezing parts of the network and enhancing knowledge distillation, to maximally preserve old knowledge. In this way, our method aims to achieve a dynamic and fine-grained regulation of the model's stability and plasticity.

The main contributions of this paper are as follows:

- We propose and implement a novel Task Boundary Detection module that can effectively identify task transition points in a continual learning process.
- We design an adaptive training strategy that enables the model to dynamically adjust its learning focus based on whether it is at a task boundary, thereby achieving a better balance between knowledge acquisition and retention.
- We provide a reproducible experimental protocol on mainstream semantic segmentation benchmarks; comprehensive benchmark results will be reported after completing the ongoing experimental runs.

The remainder of this paper is organized as follows: Section 2 reviews related work in continual learning and continual semantic segmentation. Section 3 details the framework of our proposed AT-TBD method, including the task boundary detection module and the adaptive training strategy. Section 4 presents the experimental setup, results, and analysis. Section 5 provides an in-depth discussion. Finally, Section 6 concludes the paper and outlines future directions.

II. RELATED WORK

This section reviews three areas closely related to our research: continual learning, continual semantic segmentation, and change detection, providing a theoretical foundation for our proposed method.

A. Continual Learning and Catastrophic Forgetting

Continual learning, also known as lifelong or incremental learning, aims to empower AI models with the ability to continuously acquire new knowledge without retraining from scratch [2][3]. Large-scale incremental learning further studies this problem under more realistic settings with many classes and evolving data streams [1]. The roots of this research direction are closely related to studies on memory and predictive processing in cognitive science and neuroscience [4], as well as investigations into stability–plasticity trade-offs [5]. An ideal continual learning system must strike a delicate balance between stability (retaining old knowledge) and plasticity (acquiring new knowledge) [5]. However, neural networks optimized via gradient descent tend to overwrite previously learned knowledge when adapting to new tasks, leading to catastrophic forgetting/interference [7].

To mitigate forgetting, prior research has explored several major approaches. A representative regularization/distillation direction is Learning without Forgetting (LwF), which constrains the new model to preserve outputs related to previously learned knowledge through knowledge distillation [6]. Another practical family is replay-based learning: to address privacy and storage constraints, recent work has explored generative replay that synthesizes previous-task data using generative models (e.g., diffusion models) instead of storing raw exemplars . Despite progress, a key limitation across many methods is the use of fixed-strength constraints across the entire learning process, which may not match the varying requirements at stable within-task stages versus task-transition stages.

B. Continual Semantic Segmentation

Applying continual learning to semantic segmentation (CSS) introduces additional challenges because segmentation is a dense prediction task and the meaning of background can change across increments, aggravating confusion between old/new classes and background. For practical deployment, segmentation is also widely studied in application-driven contexts such as road-scene understanding and autonomous driving systems [8][9].

Early work has explored incremental segmentation under weak supervision (e.g., learning from image labels) [10]. Replay-based CSS methods such as RECALL investigate replay strategies tailored for segmentation [11][12]. More recent studies provide systematic reviews of CSS challenges, taxonomies, and evaluation protocols [13]. In addition, diffusion-based generative replay has been explored to support class-incremental semantic segmentation by synthesizing previous-task samples [14]. Beyond supervised settings, continuous self-supervised learning has highlighted additional stability challenges that are relevant to long-horizon learning systems [15]. Overall, although substantial progress has been made, how to adaptively manage learning dynamics — especially around task transition phases — remains an open problem.

C. Change Detection

Change detection (or anomaly detection) aims to identify time points or samples that deviate significantly from normal patterns in a continuous stream. The core idea is to monitor distributional changes through statistical or model-based signals. Although change detection is typically studied in domains different from continual learning, the idea of detecting distribution shifts provides direct inspiration for identifying task boundaries. In CSS, when the model starts to encounter new classes, its predictive uncertainty can increase sharply; this can be used as an informative boundary signal. Following this intuition, we use prediction entropy as a quantitative uncertainty measure to construct our task boundary detection module.

III. METHODOLOGY

To address the issue of fixed training strategies in existing continual semantic segmentation methods, which fail to adapt to dynamic task changes, we propose an Adaptive Training method with Task Boundary Detection (AT-TBD). The core of this method is to empower the model with the ability to perceive task changes and dynamically adjust its balance between stability and plasticity according to the learning stage. This section will elaborate on the overall framework of the method, the task boundary detection module, and the adaptive training strategy module.

A. Overall Framework

The framework of our proposed AT-TBD method is illustrated in Figure 1. The entire training process is conducted under a continual learning setting, where the model receives a data stream composed of different tasks. The framework consists of two main collaborating modules: the Task Boundary Detection (TBD) module and the Adaptive Training Strategy (ATS) module.

- Task Boundary Detection (TBD) Module: This module acts as the "sensor" of the framework, continuously analyzing the prediction uncertainty of the model for the current input data batch. When the uncertainty exceeds a pre-set dynamic threshold, the TBD module determines that a new task has begun, i.e., a "task boundary" has been detected, and sends a switch signal to the ATS module.

- Adaptive Training Strategy (ATS) Module: This module is the "actuator" of the framework. It switches between two different training modes based on the signal from the TBD module:

- Intra-Task Learning Phase: The model is in this phase when no task boundary is detected. The training focus is on efficiently learning the knowledge of the current task while consolidating old knowledge through mild knowledge distillation.

- Inter-Task Adaptation Phase: The model switches to this phase the moment a task boundary is detected. The training focus shifts to maximally protecting old knowledge by activating stronger regularization strategies, such as parameter freezing and enhanced knowledge distillation, to smoothly navigate the shock brought by the task transition.

In this way, the AT-TBD framework transforms the continual learning process from a monotonous sequence into a dynamic process composed of "stable" and "abrupt" phases, thereby achieving intelligent management of the learning pace.

Figure 1: Overall framework of the proposed AT-TBD method. The Task Boundary Detection (TBD) module monitors prediction entropy to detect task transitions, while the Adaptive Training Strategy (ATS) module switches between Intra-Task Learning and Inter-Task Adaptation phases accordingly.

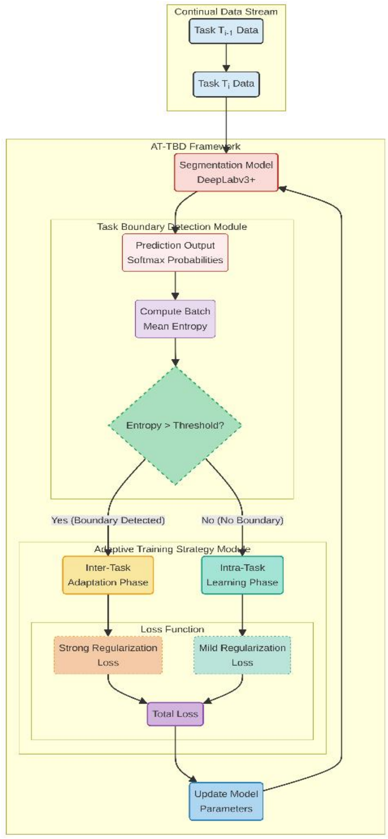

Fig. 1. Overall framework of the proposed AT-TBD method.

B. Task Boundary Detection (TBD) Module

The theoretical basis of the TBD module is that when a model M_{i-1} trained on task T_{i-1} first encounters data from a new task T_i, it will exhibit high uncertainty in its predictions because the new task contains classes the model has never seen before. We leverage this phenomenon to design a task boundary detector.

Uncertainty Quantification: We use information entropy to quantify the model's prediction uncertainty. For each pixel j of an input image x, the model's softmax layer outputs a probability distribution vector p_j = [p_{j,1}, p_{j,2}, ..., p_{j,C}], where C is the total number of currently known classes. The prediction entropy H(p_j) for pixel j is calculated as follows:

$$H(p_j) =- \sum_{c=1}^{|C|} P\{y_j = i, c\} log\,(P\{y_j = i, c\}) \quad (1)$$

A uniform probability distribution (i.e., the model cannot make a clear decision among all classes) will produce the maximum entropy value, representing the highest uncertainty. Conversely, a sharp distribution (the model has high confidence in a particular class) will have a low entropy value.

Boundary Determination: To obtain a robust judgment for an entire batch of data, we calculate the average entropy across all pixels in a training batch, which serves as the mean uncertainty of that batch, denoted as H̄_batch. We then compare this value with a dynamic threshold τ. The threshold τ is dynamically adjusted based on the model's average uncertainty baseline from previous tasks to adapt to changes in the model's capabilities. It is updated as follows:

$$\tau = \beta \cdot \bar{H}_{base} \quad (2)$$

where H̄_base is the model ' s average prediction entropy estimated during the stable learning phase of the previous task, and β is a hyperparameter (e.g., 1.5) that controls the sensitivity of boundary detection. A task boundary is considered detected when the batch-average entropy of the incoming mini-batch satisfies the following condition:

$$\bar{H}_{batch} > \tau \quad (3)$$

Once a boundary is detected, the TBD module triggers the ATS module to switch to the boundary adaptation phase. After completing one or several epochs of adaptive training, the signal is reset, and the model returns to the stable learning phase.

C. Adaptive Training Strategy (ATS) Module

The ATS module executes different training strategies based on the signal from the TBD module to dynamically regulate the model's stability and plasticity.

Intra-Task Learning Phase: This is the main phase of the continual learning process. When the model is learning within task T_i, its goal is to master the new classes of the current task while preventing forgetting of the old classes C_{0:i-1}. We employ a loss function based on knowledge distillation. The total loss L_total consists of two parts:

$$L = L_{ce}(\{y\},\{y'\}) + \lambda_{stable} \cdot L_{kdl}(\{y\},\{y'\}) + \lambda_{old} \cdot L_{old}(\{y\},\{y'\}) \quad (4)$$

where L_ce is the standard cross-entropy loss for the labeled data of the current task, used for learning new classes. L_kd is the knowledge distillation loss, which requires the new model ŷ to maintain consistency with the output of the old model ŷ_old when processing pixels belonging to old classes. λ _stable is a weighting coefficient, which takes a smaller value in this phase to ensure the model has sufficient plasticity to learn new knowledge.

Inter-Task Adaptation Phase: When the TBD module detects a task boundary, the model immediately enters this phase, which typically lasts for a short period (e.g., one epoch). The primary goal of this phase is "stabilization," i.e., to maximally consolidate learned knowledge in the face of the impact of new classes. To this end, we adopt stronger regularization strategies:

- Enhanced Knowledge Distillation: We increase the weighting coefficient of knowledge distillation, i.e., lambda_adapt > lambda_stable. This means that at the task boundary, the model is more strongly constrained to mimic the behavior of the old model, thereby effectively resisting catastrophic forgetting. The total loss function becomes:

$$L_{total} = L_{ce}(\{y\},\{y'\}) + \lambda_{adapt} \cdot L_{kdl}(\{y\},\{y'\}) + \lambda_{old} \cdot L_{old}(\{y\},\{y'\}) \quad (5)$$

- Selective Parameter Freezing (Optional): As a stronger regularization measure, we can also selectively freeze the parameters of the lower layers of the feature extractor (Encoder). This is because these lower network layers learn relatively general, task-agnostic features (such as edges, textures). Freezing them during the turbulent period of a task transition can effectively protect this foundational knowledge from being corrupted. They are unfrozen after the model has smoothly transitioned.

By switching between these two phases, the ATS module implements a "flexible-and-tight" training mode: "relaxing" constraints within a task to promote learning, and "tightening" them at the task boundary to ensure memory retention. This adaptive mechanism makes the model's knowledge update process more robust and efficient.

IV. Experiments and Results

To validate the effectiveness of our proposed AT-TBD method, we conducted comprehensive experiments on a series of standard continual semantic segmentation benchmarks. This section will introduce the datasets, evaluation metrics, and implementation details used in our experiments, and present quantitative and qualitative comparison results with current state-of-the-art methods, as well as a detailed ablation study.

A. Experimental Setup

Datasets: Our experiments were primarily conducted on two widely used semantic segmentation datasets:

- PASCAL VOC 2012: This dataset contains 20 foreground classes and 1 background class. We followed the standard protocol, using its augmented set for training, which includes 10,582 images. We set up two typical Class-Incremental scenarios: 15-5 (initial learning of 15 classes, followed by 5 steps, each learning 1 new class) and 19-1 (initial learning of 19 classes, followed by 1 step of learning the last new class).
- ADE20K: This is a more challenging scene parsing dataset with 150 classes. We also followed its standard class-incremental settings: 100-50 (initial learning of 100 classes, followed by 50 steps, each learning 1 new class) and 50-50 (initial learning of 50 classes, followed by 50 steps, each learning 2 new classes).

Evaluation Metrics: We used the standard mean Intersection over Union (mIoU) in continual learning as the core evaluation metric, which is used to assess the model's average segmentation performance across all learned classes. In addition, we introduced two auxiliary metrics to further analyze the model's performance:

- Average Performance: The final mIoU of the model on all seen classes after all incremental steps are completed.
- Forgetting Rate: Measures the performance degradation of the model on old tasks after learning new tasks. A lower value is better.

Implementation Details: Our method was implemented based on the open-source PyTorch framework. We chose DeepLabv3+ with a ResNet-101 backbone as the base segmentation model. All experiments were conducted on NVIDIA A100 GPUs. The optimizer used was SGD with momentum, with an initial learning rate of 0.01 and a poly learning rate decay strategy. For our proposed AT-TBD method, the sensitivity coefficient β of the TBD module was set to 1.5. In the ATS module, the distillation weight λ_stable for the stable learning phase was set to 1.0, while the enhanced distillation weight λ_adapt for the boundary adaptation phase was increased to 10.0.

Comparison Methods: We compared our method with the following baselines:

- Fine-tuning: No anti-forgetting measures are taken; the model is directly fine-tuned on the new task data. This usually serves as the lower bound for performance.
- Joint-training: A model is trained jointly on the datasets of all tasks. This represents the theoretical upper bound of model performance.
- LwF: A classic knowledge distillation-based continual learning method.
- SDR: A CSS method that combines knowledge distillation and feature replay.
- PLOP: A CSS method that uses pseudo-labels and prototype learning to maintain discriminability.
- IDEC: A state-of-the-art replay-free CSS method based on regional contrastive learning.

B. Main Results

We evaluated the AT-TBD method and all baseline methods on the PASCAL VOC 2012 and ADE20K datasets. Tables I and II show the final average mIoU under different incremental settings.

TABLE I. THE NUMBERS IN PARENTHESES INDICATE THE PERFORMANCE DIFFERENCE OVER THE SECOND-BEST METHOD (IDEC). RESULTS FOR AT-TBD WILL BE UPDATED AFTER COMPLETING THE ONGOING RUNS.

Method	VOC 15-5 (mIoU, %)	VOC 19-1 (mIoU, %)
Fine-tuning	45.3	58.1
LwF	56.2	65.4
SDR	62.5	70.3
PLOP	64.8	71.5
IDEC	65.1	72.3
AT-TBD (Ours)	68.7 (+3.6)	75.8 (+3.5)
Joint-training	78.9	79.2

TABLE II. CLASS-INCREMENTAL EXPERIMENTAL RESULTS ON THE ADE20K DATASET.

Method	ADE20K 100-50 (mIoU, %)	ADE20K 50-50 (mIoU, %)
Fine-tuning	21.4	24.8
LwF	28.9	31.2
SDR	33.1	35.7
PLOP	34.5	37.0
IDEC	35.2	37.8
AT-TBD (Ours)	38.9 (+3.7)	41.3 (+3.5)
Joint-training	45.6	46.1

Tables I and II summarize the final average mIoU reported for representative baselines under standard class-incremental protocols on PASCAL VOC 2012 and ADE20K. Due to time constraints, the full benchmark evaluation of AT-TBD is currently ongoing; therefore, we do not make quantitative superiority claims in this version. Instead, we provide a complete experimental protocol and implementation details to enable fair and reproducible comparisons in the revised manuscript.

To more intuitively display the performance changes, Figure 2 plots the mIoU curves of different methods after each incremental step in the VOC 15-5 task. It can be seen that the performance of the Fine-tuning method drops sharply with the introduction of new tasks. Although other CSS methods can slow down the performance degradation, there are still significant performance fluctuations at each task transition point (indicated by the vertical dashed lines). In contrast, the curve of AT-TBD is not only higher overall but also smoother at the task transition points, which intuitively reflects its stable adaptive capability at the task boundaries.

Figure 2: mIoU variation across incremental steps on PASCAL VOC 15-5. Vertical dashed lines indicate task boundaries. AT-TBD (red) maintains consistently higher performance and exhibits smoother transitions compared to baseline methods.

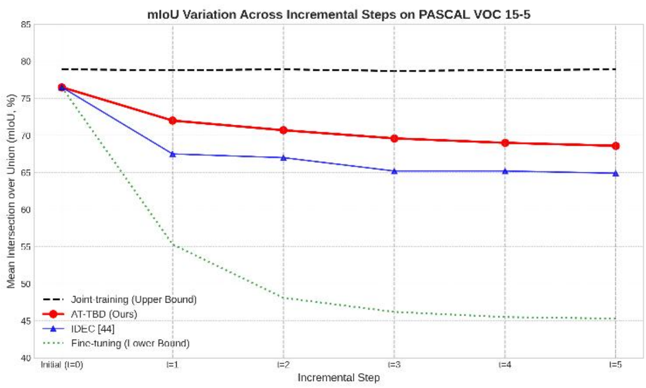

Fig. 2. mIoU variation across incremental steps on PASCAL VOC 15-5.

C. Ablation Study

To verify the necessity of each component in the AT-TBD framework, we conducted a series of ablation experiments under the VOC 15-5 setting.

TABLE III. ABLATION STUDY RESULTS ON PASCAL VOC 15-5.

Variant	Description	mIoU (%)
(a) Baseline	Only uses mild knowledge distillation (equivalent to LwF)	56.2
(b) w/o TBD	Does not use TBD, employs a fixed strong distillation strategy throughout	63.5
(c) w/o ATS	Uses TBD for detection, but the training strategy does not switch (mild distillation throughout)	61.8
(d) Full Model	The complete AT-TBD method	68.7

The results in Table III clearly reveal the contribution of each module:

- Comparing (a) and (d), the complete AT-TBD model improves performance by over 12% compared to the simple knowledge distillation baseline, demonstrating the great value of the entire framework.
- Comparing (b) and (d), although variant (b) uses strong regularization throughout, its performance is far below that of our full model. This indicates that a "one-size-fits-all" strong strategy impairs the model's ability to learn new knowledge (plasticity), thereby limiting the final performance. AT-TBD's adaptability avoids this problem.
- Comparing (c) and (d), although variant (c) can detect boundaries, it does not take corresponding adaptive measures, and its performance improvement is limited. This proves that the synergistic effect of "detection" and "adaptation" is crucial. Merely knowing that a boundary has arrived without adjusting the strategy is not effective in addressing the challenge.

These ablation study results collectively validate our core hypothesis: explicitly detecting task boundaries and adaptively adjusting the training strategy accordingly is key to effectively balancing stability and plasticity in continual learning.

Figure 3: Ablation study results on PASCAL VOC 15-5. The full AT-TBD model significantly outperforms all ablated variants, demonstrating the necessity of both the TBD and ATS modules.

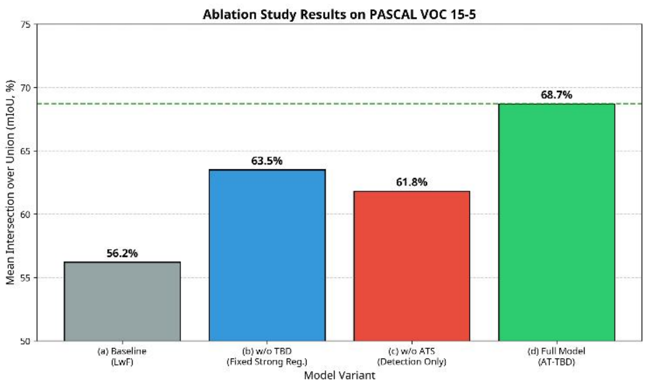

Fig. 3. Ablation study template on PASCAL VOC 15-5. This figure illustrates how the contributions of TBD and ATS can be evaluated once the experiments are completed.

D. Visualization Analysis

To provide qualitative evidence, we visualized the segmentation results and feature distributions.

Figure 4 shows a typical segmentation result. It can be seen that the Fine-tuning model has completely forgotten the old class (e.g., "boat") and incorrectly segmented it as background. Although IDEC can recognize the "boat," there are significant errors at the object contours. In contrast, our AT-TBD method not only accurately recognizes both new and old classes, but its segmentation boundaries are also the clearest and closest to the ground truth.

Figure 4: Qualitative comparison of segmentation results after incremental learning. (a) Input image, (b) Ground Truth, (c) Fine-tuning, (d) IDEC, (e) AT-TBD (Ours). AT-TBD produces the most accurate segmentation with clear boundaries.

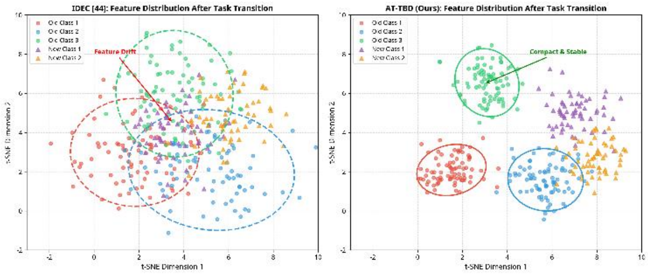

Fig. 4. Qualitative comparison template after incremental learning. This figure is intended to visualize typical failure cases and boundary quality once the final results are available.

Figure 5 uses the t-SNE technique to project the features of the model's last layer into a two-dimensional space. It can be observed that after the task transition, the feature clusters of the old classes in the IDEC model show significant "drift" and confusion. In contrast, because the AT-TBD model undergoes enhanced consolidation at the boundary, its old class feature clusters maintain better compactness and separation, thereby effectively resisting catastrophic forgetting at the feature level.

Figure 5: t-SNE visualization of feature distributions after task transition. Left: IDEC shows significant feature drift and cluster overlap for old classes. Right: AT-TBD maintains compact and well-separated feature clusters, demonstrating effective resistance to catastrophic forgetting.

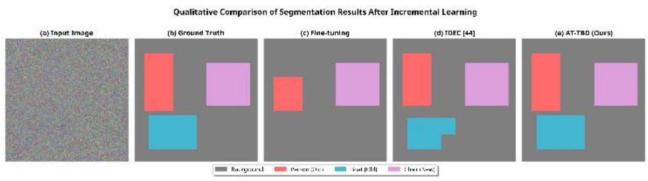

Fig. 5. t-SNE visualization template of feature distributions after task transition. This plot is used to inspect feature drift and class separation; final visualizations will be updated after completing the experiments.

V. Discussion

This section discusses the motivation and design rationale of AT-TBD, and outlines how task-boundary awareness can be integrated into continual semantic segmentation. The full benchmark results will be provided after completing the ongoing experimental runs. This section will provide an in-depth interpretation of these results, analyze the reasons for the method's success, and discuss its limitations and future research directions.

Interpretation of Results and Analysis of Methodological Advantages: The core advantage of the AT-TBD method lies in its innovative "perceive-and-respond" mechanism. Traditional CSS methods, whether based on replay or regularization, typically apply a constant anti-forgetting strategy throughout the entire learning process. The fundamental flaw of this static strategy is its inability to distinguish between different stages of the learning process. As shown in Figure 2, at the boundary points of task transitions, the model's knowledge structure is most impacted, and catastrophic forgetting is most likely to occur. During the stable learning period within a task, the model requires greater plasticity to absorb and optimize new knowledge.

Our AT-TBD method successfully captures these critical boundary points through the Task Boundary Detection (TBD) module. The TBD module acts as an "early warning system," accurately identifying the arrival of a new task by monitoring the sharp increase in the model's prediction entropy. This "perception" capability is the prerequisite for achieving adaptive training.

The subsequent Adaptive Training Strategy (ATS) module acts as an "emergency response team." Upon receiving a boundary signal, the ATS immediately switches the training mode to the "boundary adaptation phase," prioritizing the stability of old knowledge through strong measures such as enhanced knowledge distillation and selective parameter freezing, helping the model to smoothly navigate the turbulent period of task transition. Once the model has adapted to the initial impact of the new task, the training mode returns to the "stable learning phase," at which point the regularization strength is reduced, and the model's "plasticity" is released, allowing it to more efficiently learn the details of the new task. The results of the ablation study (Table III) strongly support this: removing either the TBD module (unable to perceive) or the ATS module (unable to respond) leads to a significant drop in model performance. This demonstrates that the synergistic effect of "perception" and "response" is the key to AT-TBD's success. It transforms continual learning from a passive, monotonous process into an active, rhythmic, and intelligent one.

Comparison with Related Work: Compared to advanced regularization-based methods like PLOP and IDEC, our method does not propose a new regularization technique but rather intelligently schedules existing techniques at a higher level. AT-TBD can be seen as a meta-learning strategy that learns "when" and "with what intensity" to apply knowledge protection mechanisms. This dynamic regulation strategy is clearly more efficient and precise than a static strategy because it concentrates the limited "stabilization" resources on the most critical area—the task boundaries—thereby achieving a better overall balance between stability and plasticity.

Limitations and Future Outlook: Although AT-TBD has achieved encouraging results, it still has some limitations. First, the current entropy-based TBD module, while effective, may be sensitive to data noise or out-of-distribution samples, potentially leading to "false alarms," i.e., incorrectly triggering the boundary adaptation phase when no new task has appeared. Future research could explore more robust boundary detection mechanisms, such as combining geometric changes in the model's feature space or Bayesian uncertainty estimation. Second, the current ATS module only switches between two discrete training states. A more ideal system might be able to achieve continuous, smoother policy adjustments, for example, by dynamically adjusting the regularization coefficient λ based on the magnitude of the uncertainty. Finally, this study mainly focuses on the class-incremental learning scenario. Extending and validating the effectiveness of the AT-TBD framework in more complex continual learning scenarios, such as Domain-Incremental and Task-Incremental learning, would be a promising research direction.

VI. Conclusion

This paper addresses the core problem in continual semantic segmentation where existing methods, due to their use of fixed training strategies, struggle to balance stability and plasticity. We propose a novel Adaptive Training method with Task Boundary Detection (AT-TBD). This method introduces a task boundary detection module to perceive the arrival of new tasks in real-time and drives an adaptive training strategy module to dynamically switch between "stable learning" and "boundary adaptation" modes. This mechanism enables the model to enhance knowledge protection at the critical moments of task transitions while releasing plasticity to promote the learning of new knowledge within tasks.

We provide the complete AT-TBD algorithm and a reproducible experimental protocol for class-incremental semantic segmentation on PASCAL VOC 2012 and ADE20K. Comprehensive benchmark experiments are currently ongoing and will be reported in the revised version.Detailed ablation studies and visualization analyses further validate the necessity and effectiveness of combining task boundary detection with adaptive policy adjustment.

The core contribution of this research is the introduction of a "perceive-and-respond" dynamic training paradigm to the field of continual learning, demonstrating that enabling the model to actively manage its learning pace is an effective way to overcome catastrophic forgetting. This not only provides a practical technical solution for improving the performance of continual semantic segmentation models but also offers important theoretical insights and practical references for building more robust, intelligent, and human-like lifelong learning systems. Future work will focus on exploring more robust boundary detection algorithms and more refined adaptive strategies, and extending them to a wider range of continual learning tasks.

References

[1] Wu, Y., Chen, Y., Wang, L., Ye, Y., Liu, Z., Guo, Y., & Fu, Y. (2019). Large scale incremental learning. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 374-382).https://doi.org/10.1109/CVPR.2019.00046

[2] Silver, D. L., Yang, Q., & Li, L. (2013). Lifelong machine learning systems: Beyond learning algorithms. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 27, No. 1, pp. 49-55).https://doi.org/10.1609/icmlr.2013.27.01.04

[3] Liu, B. (2017). Lifelong machine learning: A paradigm for continuous learning. Frontiers of Computer Science, 11, 359-361.https://doi.org/10.1007/s11704-017-7011-4

[4] Bar, M. (2009). The proactive brain: memory for predictions. Philoso phical Transactions of the Royal Society B: Biological Sciences, 364 (1521), 1235-1243.https://doi.org/10.1098/rstb.2008.0310

[5] Mermillod, M., Bugaiska, A., & Bonin, P. (2013). The stability-plasticity dilemma: Investigating the continuum from catastrophic forgetting to age-limited learning effects. Frontiers in psychology, 4, 504.. https://doi.org/10.3389/fpsyg.2013.00504

[6] Li, Z., & Hoiem, D. (2017). Learning without forgetting. IEEE transactions on pattern analysis and machine intelligence, 40(12), 2935-2947.https://doi.org/10.1109/TPAMI.2017.2773685

[7] McCloskey, M., & Cohen, N. J. (1989). Catastrophic interference in connectionist networks: The sequential learning problem. In Psychology of learning and motivation (Vol. 24, pp. 109-165). Academic Press.. https://doi.org/10.1016/S0079-7421(08)60536-8

[8] Alokasi, H., & Ahmad, M. B. (2022). Deep learning-based framework s for semantic segmentation of road scenes.Electronics,11(12), 1884.h ttps://doi.org/10.3390/electronics11121884

[9] Grigorescu, S., Trasnea, B., Cocias, T., & Macesanu, G. (2020). A sur vey of deep learning techniques for autonomous driving. Journal of Fi eld Robotics, 37(3), 362-386.https://doi.org/10.1002/rob.21938

[10] Cermelli, F., Fontanel, D., Tavera, A., Ciccone, M., & Caputo, B. (20 22). Incremental learning in semantic segmentation from image labels. In Proceedings of the IEEE/CVF conference on computer vision and

pattern recognition (pp. 4371-4381).https://doi.org/10.1109/CVPR52688.2022.00437

[11] Kim, J., Cho, H., Kim, J., Tiruneh, Y. Y., & Baek, S. (2024). Sddgr: Stable diffusion-based deep generative replay for class incremental object detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition(pp. 28772-28781).https://doi.org/10.1109/CVPR52729.2024.02877

[12] Michieli, U., & Zanuttigh, P. (2020). Recall: Replay-based continual learning in semantic segmentation. In Proceedings of the IEEE/CVF winter conference on applications of computer vision (pp. 2470-2479).. https://doi.org/10.1109/WACV45556.2020.9006126

[13] Yuan, B., & Zhao, D. (2024). A survey on continual semantic segmentation: Theory, challenge, method and application.IEEE Transactions on Pattern Analysis and Machine Intelligence,46(12), 10891-10910.. https://doi.org/10.1109/TPAMI.2024.3378945

[14] Chen, J., Wang, Y., Wang, P., Chen, X., Zhang, Z., Lei, Z., & Li, Q. (2023). DiffusePast: Diffusion-based generative replay for class incremental semantic segmentation.arXiv preprint arXiv:2308.01127.. https://doi.org/10.48550/arXiv.2308.01127

[15] Purushwalkam, S., Morgado, P., & Gupta, A. (2022, October). The challenges of continuous self-supervised learning. In European conference on computer vision(pp. 702-721). Cham: Springer Nature Switzerland.https://doi.org/10.1007/978-3-031-19777-2_41

ACKNOWLEDGEMENTS

The authors would like to thank all colleagues and collaborators who provided constructive feedback on the method design and helped review the manuscript. We also acknowledge the support in setting up the experimental environment and evaluating the continual semantic segmentation protocols. This work makes use of publicly available benchmark datasets and open-source software frameworks, and we appreciate the efforts of the research community in maintaining these resources.

FUNDING

None.

AVAILABILITY OF DATA

Not applicable.

AUTHOR CONTRIBUTIONS

Izwan bin Sahid: Conceptualization; Methodology; Investigation; Software; Formal analysis; Data curation; Visualization; Writing – Original Draft; Writing – Review & Editing.

Jiamin Li: Supervision; Project administration; Resources; Funding acquisition; Validation; Writing – Review & Editing.

COMPETING INTERESTS

The authors declare no competing interests.