Distribution-Aware Continual Test Time Adaptation for Semantic Segmentation

Jiayi Ni*, Senqiao Yang*, Jiaming Liu†, Xiaoqi Li, Wenyu Jiao, Ran Xu,
Zehui Chen, Yi Liu, Shanghang Zhang ✉

Abstract— Since autonomous driving systems usually face dynamic and ever-changing environments, continual test-time adaptation (CTTA) has been proposed as a strategy for transferring deployed models to continually changing target domains. However, the pursuit of long-term adaptation often introduces catastrophic forgetting and error accumulation problems, which impede the practical implementation of CTTA in the real world. Recently, existing CTTA methods mainly focus on utilizing a majority of parameters to fit target domain knowledge through self-training. Unfortunately, these approaches often amplify the challenge of error accumulation due to noisy pseudo-labels, and pose practical limitations stemming from the heavy computational costs associated with entire model updates. In this paper, we propose a distribution-aware tuning (DAT) method to make the semantic segmentation CTTA efficient and practical in real-world applications. DAT adaptively selects and updates two small groups of trainable parameters based on data distribution during the continual adaptation process, including domain-specific parameters (DSP) and task-relevant parameters (TRP). Specifically, DSP exhibits sensitivity to outputs with substantial distribution shifts, effectively mitigating the problem of error accumulation. In contrast, TRP are allocated to positions that are responsive to outputs with minor distribution shifts, which are fine-tuned to avoid the catastrophic forgetting problem. In addition, since CTTA is a temporal task, we introduce the Parameter Accumulation Update (PAU) strategy to collect the updated DSP and TRP in target domain sequences. We conduct extensive experiments on two widely-used semantic segmentation CTTA benchmarks, achieving promising performance compared to previous state-of-the-art methods.

I. INTRODUCTION

Continuous semantic segmentation represents a fundamental task in scene understanding, particularly in autonomous driving [1], [2]. Assuming that both training and test data are under the same distribution, deployed models can achieve remarkable performance. However, in the real world, this assumption is frequently violated due to the continually changing environments, resulting in a significant degradation in performance [3], [4]. Therefore, the concept of continual test-time adaptation (CTTA) has been introduced [5]. Unlike traditional test-time adaptation, CTTA focuses on handling a sequence of distinct distribution shifts over time, rather than just a single shift. In addition, due to data privacy, the CTTA process renders the source data inaccessible, making it more challenging but expanding its applicability [6].

Jiaming Liu, Jiayi Ni, Senqiao Yang, Xiaoqi Li, Ran Xu, Zehui Chen and Shanghang Zhang are with National Key Laboratory for Multimedia Information Processing, School of CS, Peking University. Wenyu Jiao is with University of Washington. Yi Liu is with Baidu. *: Equal Contribution. †: Project Leader. ✉ Corresponding Author: shanghang@pku.edu.cn

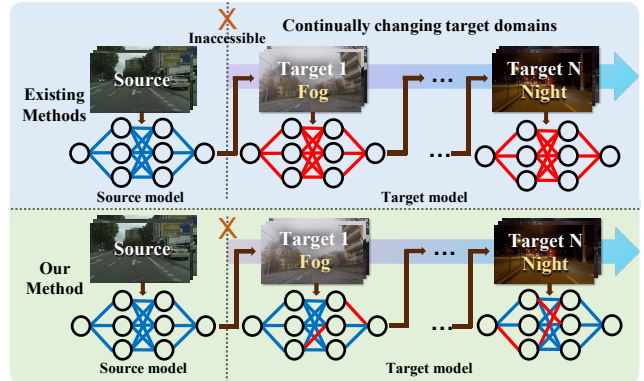Our project will be released at https://github.com/RochelleNi/DAT

Fig. 1. The CTTA problem aims to adapt the source model to continually changing target domains. Existing model-based methods focus on utilizing a majority of parameters to fit different target domain knowledge in continually changing environments. Differently, we propose a parameter-efficient model-based method named distribution-aware tuning (DAT), a novel paradigm for stable continual adaptation. DAT adaptively selects two small groups (e.g., 5%) of trainable parameters based on the data distribution, enabling the simultaneous extraction of domain-specific and task-relevant knowledge. Instead of updating the entire model, DAT streamlines the continual adaptation process, enhancing its efficiency and potential for real-world applications. The red lines indicate the updated parameters.

The pursuit of continual adaptation often introduces challenges such as catastrophic forgetting and error accumulation, hindering the practical implementation of CTTA in real-world applications. To address these issues, existing CTTA approaches have primarily employed model-based methods [5], [7] or parameter-efficient methods [8], [9] to simultaneously extract target domain-specific and domain-invariant knowledge. As shown in Fig .1, the conventional model-based methods tend to update a majority of parameters for learning target domain knowledge through self-training. Unfortunately, these approaches frequently exacerbate the problem of error accumulation by introducing noisy pseudo-labels. Additionally, they pose practical limitations due to the substantial computational costs associated with updating the entire model. Meanwhile, parameter-efficient methods leverage soft prompts [10] or adapters [11] with limited trainable parameters to learn long-term domain-shared knowledge, leading to difficulties in preventing catastrophic forgetting.

In this paper, we propose a parameter-efficient model-based method named distribution-aware tuning (DAT), a novel paradigm for stable continual adaptation, shown in Fig .1. DAT does not require the addition of any prompts or adapters, avoiding disruption to the model's original structure. Instead, it leverages two small groups (around 5%) of its own parameters to extract target domain-specific and task-relevant knowledge simultaneously. Specifically, DAT first evaluates

the degree of pixel-level distribution shift for each model output using the uncertainty mechanism [12], [13], [14]. At the same time, it divides the majority of pixels into two groups, representing significant distribution shifts and minor distribution shifts, respectively. DAT fine-tunes the selected domain-specific parameters (DSP) that exhibit sensitivity to outputs with substantial distribution shifts, effectively mitigating the problem of error accumulation. In contrast, task-relevant parameters (TRP) are allocated to positions that are responsive to outputs with minor distribution shifts, which are fine-tuned to avoid the catastrophic forgetting problem.

Regarding the process of DSP and TRP selection, since CTTA is a task involving temporal sequences, we introduce the Parameter Accumulation Update (PAU) strategy to collect parameters. When processing a series of target domain samples, we only select a very small fraction of parameters (e.g., 0.1%) for each sample to update and add these parameters to the parameter group. This process continues until the uncertainty value becomes relatively small, indicating that the collected parameter group can better understand current target domains. After the accumulation stage, the accumulated parameter group (e.g., 5.0%) will continue to be updated on the subsequent sequence data until encountering the next changing environment.

The main contributions can be outlined as follows:

1) We introduce an efficient distribution-aware tuning (DAT) method that adaptively selects two small groups of trainable parameters (e.g., 5.0%), domain-specific parameters (DSP) and task-relevant parameters (TRP), based on the degree of distribution shifts. This approach addresses issues related to error accumulation and catastrophic forgetting.

2) For DSP and TRP selection, we employ the parameter accumulation update (PAU) strategy. When processing early target domain samples within the sequence, we select only a very small fraction of parameters (e.g., 0.1%) for each sample and add these parameters to the updated parameter group until the distribution shift becomes relatively small.

3) Our proposed method achieves competitive performance compared to previous state-of-the-art methods on two CTTA benchmarks, Cityscape-ACDC [4] and SHIFT [3] continual datasets, showcasing its effectiveness in addressing the semantic segmentation CTTA problem.

II. RELATED WORK

Test-time adaptation (TTA), also known as source-free domain adaptation [6], aims to adapt a source model to an unknown target domain distribution without the utilization of any source domain data. These techniques mainly focus on unsupervised methods, such as self-training and entropy regularization, for the meticulous fine-tuning of the source model. Specifically, SHOT [15] distinguishes itself by concentrating on optimizing the feature extractor exclusively through information maximization and pseudo-labeling. AdaContrast [16], while also leveraging pseudo labeling for TTA, incorporating self-supervised contrastive learning techniques. Regarding model-based adaptation strategies, [17] introduces a methodology involving the adjustment

of the output distribution, showcasing its effectiveness in mitigating domain shift issues. Tent [18] updates the training parameters in the batch normalization layers through entropy minimization. This approach has garnered attention and has prompted subsequent exploration in recent works [19], [20], which continue to investigate the robustness of normalization layers. While the aforementioned works primarily focus on classification TTA tasks, there has been a recent surge of interest in applying TTA to dense prediction tasks [21], [22].

Continual Test-Time Adaptation (CTTA) is a scenario where the target domain is dynamic, presenting increased challenges for traditional TTA methods. Existing CTTA approaches have predominantly utilized either model-based methods [5], [7] or parameter-efficient methods [8], [9] to concurrently address error accumulation and catastrophic forgetting issues. For model-based methods, CoTTA [5] enhances accuracy by employing weighted averaging and improved average prediction, all the while preventing catastrophic forgetting through random parameter recovery. Ecotta [7], on the other hand, introduces a meta-network to combat error accumulation by regularizing the outputs from both the meta-network and the frozen network. Conversely, RMT [23] addresses error accumulation through gradient analysis and introduces a symmetric cross-entropy loss tailored for CTTA. Additionally, [24] combines self-training from the supervised branch with pseudo labels from the self-supervised branch, providing a solution for the depth estimation CTTA problem. For parameter-efficient methods, [8] introduces the use of visual domain prompts specifically at the input level, extracting domain-specific and domain-agnostic knowledge. SVDP [25] suggests the implementation of sparse visual prompts to address occlusion problems that arise from dense image-level prompts [8], thereby extracting sufficient local domain-specific knowledge. Lastly, [9] proposes ViDA to address error accumulation and catastrophic forgetting by leveraging low-rank and high-rank adapters to present different domain representations. In this paper, our focus is on addressing the semantic segmentation CTTA problem, where we propose a parameter-efficient model-based method. Our aim is to utilize a small portion of its own parameters for a stable and efficient continual adaptation, without the need for additional prompts or adapters.

III. METHOD

A. Preliminary

Continual Test-Time Adaptation (CTTA) aims to adapt the source pre-trained model to constantly changing target domains during inference. A pre-trained model, $f_\theta(x)$, with parameters θ trained on source data $\mathscr{X}_S$ and labels $\mathscr{Y}_S$, adapts online, refining predictions in a continually evolving target data without revisiting source data. The distributions of the target domains (i.e., $\mathscr{X}_{T_1}, \mathscr{X}_{T_2}, \ldots, \mathscr{X}_{T_n}$) are in a state of constant change, and each sample within these target domains can only be utilized once.
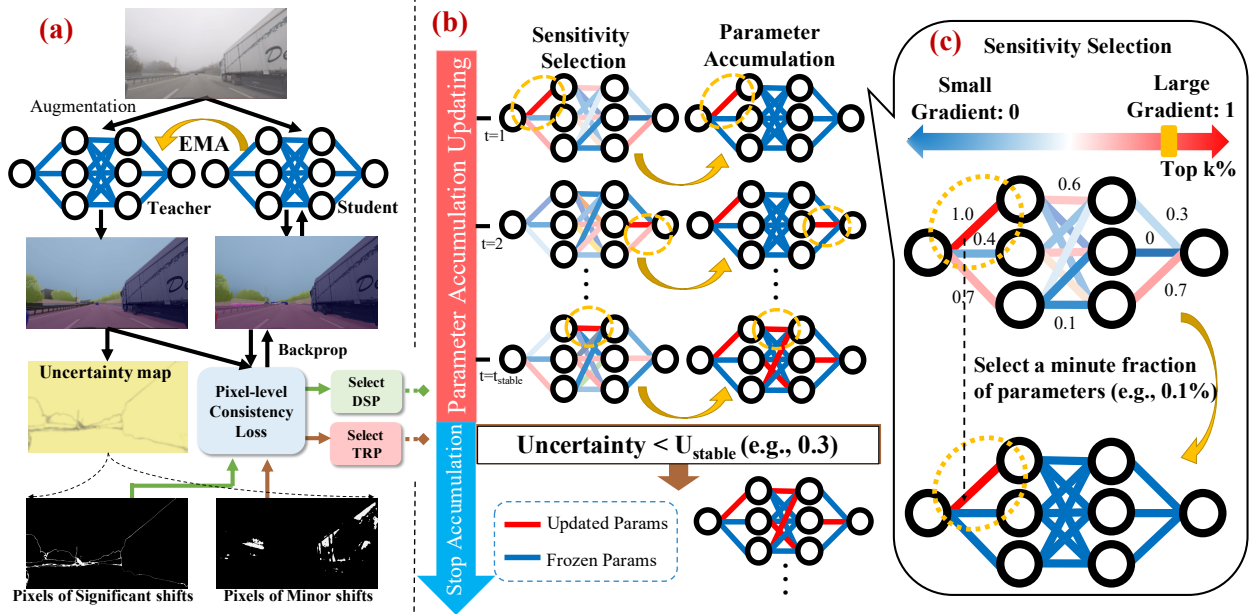
Fig. 2. **The overall framework of distribution-aware tuning.** (a) In our approach, we implement a teacher-student framework to predict semantic segmentation maps, utilizing a pixel-wise consistency loss for optimization. Simultaneously, we obtain an uncertainty map from the teacher model to evaluate pixel-level distribution shifts. Based on the degree of these shifts, we selectively choose two small groups of trainable parameters (e.g., 5.0%), namely domain-specific parameters (DSP) and task-relevant parameters (TRP). (b) In the process of DSP and TRP selection, we employ the Parameter Accumulation Update (PAU) strategy to gather these parameters. When processing a series of target domain samples, we opt to select and update only a minute fraction of parameters (e.g., 0.1%) for each sample, subsequently adding these parameters to the selected parameter group. (c) We show the details of DSP and TRP selection, which involves identifying the most sensitive parameters for pixels with significant and minor distribution shifts, respectively.

B. Overall framework and Motivation

The overall framework, as illustrated in Figure 2 (a), begins with a teacher-student framework for predicting semantic segmentation maps, penalized by pixel-wise consistency loss. Simultaneously, we obtain an uncertainty map from the teacher model to assess pixel-level distribution shifts. Based on the uncertainty values, we divide all pixels from the student segmentation map into two groups: those with significant distribution shifts and those with minor shifts. We then use the same consistency loss to measure parameter gradient sensitivity for both groups, selecting the most sensitive parameters to represent domain-specific parameters (DSP) and task-relevant parameters (TRP), as depicted in Figure 2 (c). The Parameter Accumulation Update (PAU) strategy, shown in Figure 2 (b), involves selecting a very small fraction of parameters (e.g., 0.1%) for each sample and adding these parameters to the updated parameter group (represented by red lines). This process continues until the uncertainty value becomes relatively small, indicating that the parameter group is better suited to understand the current target domains. Note that, both DSP and TRP are selected through the PAU strategy during the continual adaptation process.

Teacher-Student framework. Drawing from the insight that the mean teacher predictions often outperform standard models [26], we utilize a teacher model to enhance the accuracy of pseudo labels during continual domain adaptation. Furthermore, recognizing the robustness of the teacher-student framework in dynamic environments [23], we employ it to ensure stability within the target domains.

Distribution-aware tuning. Different from previous model-based CTTA methods, which update the majority of parameters for learning target domain knowledge, our approach aims to establish a parameter-efficient paradigm for stable continual adaptation. We achieve this by adaptively selecting two small groups of trainable parameters based on data distribution, allowing us to extract both target domain-specific and task-relevant knowledge. This approach allows us to update only the most crucial parameters for target domain samples, rather than updating the entire model, significantly enhancing the efficiency and practicality of CTTA in real-world applications. Different from previous parameter-efficient CTTA methods, our approach does not require the additional module of prompts or adapters, preserving the model's original structure. Instead, we simply fine-tune two small groups of its existing parameters (DSP and TRP) to simultaneously avoid catastrophic forgetting and error accumulation problems.

C. Distribution Aware Tuning

Uncertainty Scheme is an effective method for quantifying the degree of the model's distribution shift [9], [13], [14]. While in previous works, the confidence score is a common measure to assess prediction reliability, it tends to fluctuate irregularly and becomes unreliable in scenarios characterized by distribution shifts. To overcome this limitation, we employ the MC Dropout [27] on the prediction head, allowing multiple forward propagations to obtain sets of segmentation maps (m sets) for each sample. We compute the uncertainty value for a given input. This value is defined as follows:

$$\mathscr{U}(\widetilde{x}_j) = \left(\frac{1}{m} \sum_{i=1}^{m} \| p_i(\widetilde{y}_j | \widetilde{x}_j) - \mu \|^2 \right)^{\frac{1}{2}} \qquad (1)$$

Where $p_i(\widetilde{y}_j|\widetilde{x}_j)$ denotes the predicted probability of output $\widetilde{y}_j$ based on the input pixel $\widetilde{x}_j$ during the ith iteration of forward propagation; μ signifies the average prediction over m iterations for $\widetilde{x}_j$. Therefore, $\mathscr{U}(\widetilde{x}_j)$ represents the uncertainty regarding pixel-wise model input.

Hence, we can utilize the uncertainty value to classify the pixels of the output. The formulation is as follows:

$$\begin{cases} \widetilde{x}_j \in \mathscr{G}_h, & \mathscr{U}(\widetilde{x}_j) \geq \Theta_h \\ \widetilde{x}_j \in \mathscr{G}_l, & \mathscr{U}(\widetilde{x}_j) < \Theta_l \end{cases} \quad (2)$$

The threshold value of uncertainty is denoted as Θ, where $\Theta_h = 0.6$, $\Theta_l = 0.25$. $\mathscr{G}_h$ and $\mathscr{G}_l$ represent the pixel groups with high and low uncertainty value.

Domain-Specific Parameters (DSP). After calculating the uncertainty of each pixel, we select those with significant uncertainty values, indicating substantial distribution shifts. Subsequently, we employ these pixels to select domain-specific parameters. This is specifically achieved by utilizing a binary mask to compute the loss and perform backpropagation exclusively for these pixels. Following this, we choose parameters with higher gradient values as Domain-Specific parameters. These domain-specific parameters exhibit sensitivity to outputs with substantial distribution shifts. By updating these domain-specific parameters, they effectively capture domain-specific knowledge in the target domain, thus efficiently mitigating the issue of error accumulation.

Task-Relevant Parameters (TRP). Apart from fine-tuning domain-specific parameters (DSP) to prevent error accumulation, task-relevant parameters (TRP) are selected to mitigate catastrophic forgetting. Using the identical procedure as described previously, we select pixels with lower uncertainty values, signifying minimal domain shifts and yielding more dependable segmentation outcomes. Similarly, we only backpropagate through these pixels with lower uncertainty and select the parameters with the highest gradient as Task-Relevant Parameters, which are more focused on the task itself and are not affected by domain shifts.

Overall, based on the distribution of each sample, we select domain-specific parameters and task-relevant parameters. Then, we add them in the selected parameter group and utilize the PAU method to continually update these parameters.

D. Parameter Accumulation Update

For the basic parameter update strategy, we employ Eq .4 to update the DSP and TRP of student model ($\mathscr{S}_{target}$) and leverage the widely-used exponential moving average (EMA) to update the corresponding parameters of teacher model ($\mathscr{T}_{mean}$), ensuring maximal model plasticity. The EMA process can be formulated as:

$$\mathscr{T}_{mean}^t = \alpha \mathscr{T}_{mean}^{t-1} + (1-\alpha)\mathscr{S}_{target}^t \quad (3)$$

When $t = 0$ (representing the initial time step), we employ the pre-trained model from the source domain to initialize the weights of both the teacher and student models. Additionally, we set α to 0.999, as referenced in [28].

Regarding the process of DSP and TRP selection, since CTTA and autonomous driving systems are both tasks involving temporal sequences, we introduce the Parameter Accumulation Update (PAU) strategy in addition to the existing EMA parameter updating approach. Specifically, when processing a series of target domain samples, for each individual image, we select Domain-Specific parameters and Task-Relevant Parameters based on their uncertainty. It's important to note that, to ensure the efficiency of the model, we only choose a very small fraction of parameters (e.g., 0.1%) for each sample to update, and we add these selected parameters to the parameter group. This iterative process continues until the uncertainty value becomes relatively small, indicating that the accumulated parameter group better comprehends the current target domain. After the accumulation stage, the accumulated parameter group will continue to be updated on the subsequent sequence data. When encountering the next target domain, we reset all accumulated positions of DSP and TRP, starting a new round of parameter accumulation.

E. Loss Functions

Following prior CTTA research [5], the optimization objective is a consistency loss to minimize the discrepancy between outputs of teacher and student model, which is a cross-entropy loss calculated at the pixel level:

$$\mathscr{L}_{con}(\widetilde{x}) = -\frac{1}{H \times W} \sum_{w,h}^{W,H} \sum_{c}^{C} \widetilde{y}_t(w,h,c) \log \hat{y}_t(w,h,c) \quad (4)$$

In this equation, $\widetilde{x}$ is input, $\widetilde{y}_t$ is pseudo labels generated by the teacher model and $\hat{y}_t$ is student model output, whereas C denotes the number of categories, and H and W refer to the height and width of images, respectively.

IV. EXPERIMENTS

In Sec IV-A, we provide the details of the task settings for continual test-time adaptation (CTTA), as well as a description of the datasets. In Sec IV-B, we compare our method with other baselines [29], [30], [5], [31] on two commonly used CTTA benchmarks. Comprehensive ablation studies are conducted in Sec IV-C, which investigate the impact of each component.

A. Datasets and Implementation Details

Cityscapes-to-ACDC refers to a semantic segmentation task specifically designed for cross-domain learning in autonomous driving. The source model used is an off-the-shelf pre-trained segmentation model originally trained on the Cityscapes dataset [32]. The ACDC dataset [4] contains images collected in four different, previously unseen visual conditions: Fog, Night, Rain, and Snow. In the CTTA scenario, we replicate the sequence of target domains (Fog→Night→Rain→Snow) multiple times to simulate environmental changes in real-life driving scenarios.

SHIFT is an expansive synthetic driving video dataset that simulates real-world autonomous driving scenarios [3]. This dataset consists of two parts: a discrete set containing 4,250 sequences and a continuous set with an additional 600 sequences. In this configuration, the source model pre-trains

PERFORMANCE COMPARISON FOR CITYSCAPE-TO-ACDC CTTA. DURING TESTING, WE SEQUENTIALLY EVALUATE THE FOUR TARGET DOMAINS THREE TIMES. MEAN DENOTES THE AVERAGE mIoU SCORE. GAIN QUANTIFIES THE METHOD'S IMPROVEMENT COMPARED TO THE SOURCE MODEL.

	Time	t															Mean↑	Gain
	Round			1					2					3				
Method	REF	Fog	Night	Rain	Snow	Mean↑	Fog	Night	Rain	Snow	Mean↑	Fog	Night	Rain	Snow	Mean↑		
Source	NIPS2021 [29]	69.1	40.3	59.7	57.8	56.7	69.1	40.3	59.7	57.8	56.7	69.1	40.3	59.7	57.8	56.7	56.7	/
TENT	ICLR2021 [30]	69.0	40.2	60.1	57.3	56.7	68.3	39.0	60.1	56.3	55.9	67.5	37.8	59.6	55.0	55.0	55.7	-1.0
CoTTA	CVPR2022[5]	70.9	41.2	62.4	59.7	58.6	70.9	41.1	62.6	59.7	58.6	70.9	41.0	62.7	59.7	58.6	58.6	+1.9
DePT	ICLR2023[31]	71.0	40.8	58.2	56.8	56.5	68.2	40.0	55.4	53.7	54.3	66.4	38.0	47.3	47.2	49.7	53.4	-3.3
VDP	AAAI2023[8]	70.5	41.1	62.1	59.5	58.3	70.4	41.1	62.2	59.4	58.2	70.4	41.0	62.2	59.4	58.2	58.2	+1.5
DAT	**ours**	**71.7**	**44.4**	**65.4**	**62.9**	**61.1**	**71.6**	**45.2**	**63.7**	**63.3**	**61.0**	**70.6**	**44.2**	**63.0**	**62.8**	**60.2**	**60.8**	**+4.1**

on the discrete datasets and tests on the continuous validation dataset. The continuous set is designed to simulate gradual transitions between different environments, simulating real-world scenarios such as the continual transition from clear daytime conditions to nighttime in a validation sequence.

Implementation Details. We follow the basic implementation details [5], [8] to set up our semantic segmentation experiments. Specifically, in the Cityscapes-to-ACDC scenario, we employ the Segformer-B5 [29] as our segmentation model and down-sample the original image size of 1920x1080 to 960x540, which serves as network input. For the SHIFT dataset, we follow the setup details in [3] and also employ the Segformer-B5. And we downsample the original image size to 800x500 during CTTA. For both of these two datasets, we evaluate our predictions under the original resolution. We use the Adam optimizer [33] with $(\beta_1, \beta_2) = (0.9, 0.999)$, a learning rate of 3e-4 and batch size 1 for all experiments. Besides, We apply a range of image resolution scale factors [0.5, 0.75, 1.0, 1.25, 1.5, 1.75, 2.0] for the augmentation method in the teacher model. No more than 10% parameters of the model are updated in each scenario. All experiments are conducted on NVIDIA V100 GPUs.

B. Effectiveness

Cityscapes-to-ACDC. Given the source model pre-trained on Cityscapes, we conduct CTTA on ACDC dataset, which contains four severe weather that occur sequentially during the test time. As shown in Table I, the source model's mean Intersection over Union (mIoU) is only 56.7% when directly testing the source model on target domains in three rounds. In contrast, our method consistently outperforms all previous methods, delivering a 4.1% and 2.6% improvement over the source model and the previous SOTA method CoTTA, respectively. Unlike TENT [30], which updates only the normalization layer, our DAT selects task-relevant parameters for fine-tuning, strategically allocating them to positions responsive to outputs with minor distribution shifts. Notably, TENT exhibits significant catastrophic forgetting, with mIoU dropping from 56.7 to 55.0. In contrast, our DAT does not exhibit this issue, underscoring the effectiveness of fine-tuning task-relevant parameters (TRP) in mitigating catastrophic forgetting problems. In comparison to CoTTA [5], the previous sota method, our method has significant improvement across most domains. This showcases our method's capacity to extract domain knowledge by fine-tuning domain-specific

PERFORMANCE COMPARISON FOR SHIFT DATASET'S CONTINUOUS VALIDATION SET. ACC DENOTES THE AVERAGE ACCURACY SCORE.

Scenarios	*Daytime − Night*		*Clear − Foggy*		*Clear − Rainy*		Mean
Method	mIoU	ACC	mIoU	ACC	mIoU	ACC	mIoU
Source	64.22	71.84	61.51	69.87	66.14	73.87	64.01
TENT	64.21	71.67	61.65	69.76	66.11	73.68	63.99
CoTTA	68.06	74.18	64.78	71.41	69.92	76.29	67.61
Ours	**68.63**	**75.34**	**65.31**	**72.68**	**70.56**	**77.34**	**68.17**

parameters, enabling rapid adaptation to target domains, without fine-tuning redundant parameters. In conclusion, our DAT can effectively extract diverse domain knowledge while concurrently avoiding catastrophic forgetting simultaneously. Furthermore, we present the qualitative results in Fig . 3. Our method correctly distinguishes the sidewalk from the road.

SHIFT. In order to verify the generalization of our proposed method, we conduct experiments on the autonomous driving CTTA datasets, specifically the SHIFT dataset [3]. Using a source model pre-trained on SHIFT's discrete datasets, we conduct CTTA experiment on the gradually evolving SHIFT continuous validation set. As presented in Table II, the mIoU consistently demonstrates our method's superior performance compared to the state-of-the-art (SOTA) technique across all three scenarios. Notably, in the most challenging scenario—transitioning from clear to foggy conditions—our approach exhibited a significant enhancement, with mIoU increasing from 64.78 to 65.31 and mACC improving from 71.41 to 72.68. It's worth noting that although our approach has shown improvement compared to the previous method, the short length of each sequence in the SHIFT dataset and the strong dependency between the data's each frame make it challenging to fully demonstrate our method's advantage in preventing catastrophic forgetting.

C. Ablation Study

In this subsection, we evaluate the contribution of each component in the Cityscape-to-ACDC scenario.

Effectiveness of each component As presented in Tab. III Ex_2, Teacher-student (TS) structure is a common technique in TTA [5], [8], [34]. This technique facilitates the generation of pseudo-labels in the target domain. However, it manifests a mere increase of 1.0% in mIoU without the integration of our method. This serves to substantiate that the performance improvement is not merely a derivative of employing this widespread scheme. In Ex_3, by introducing fine-tuning the domain-specific parameter (DSP), we observe that the mIoU
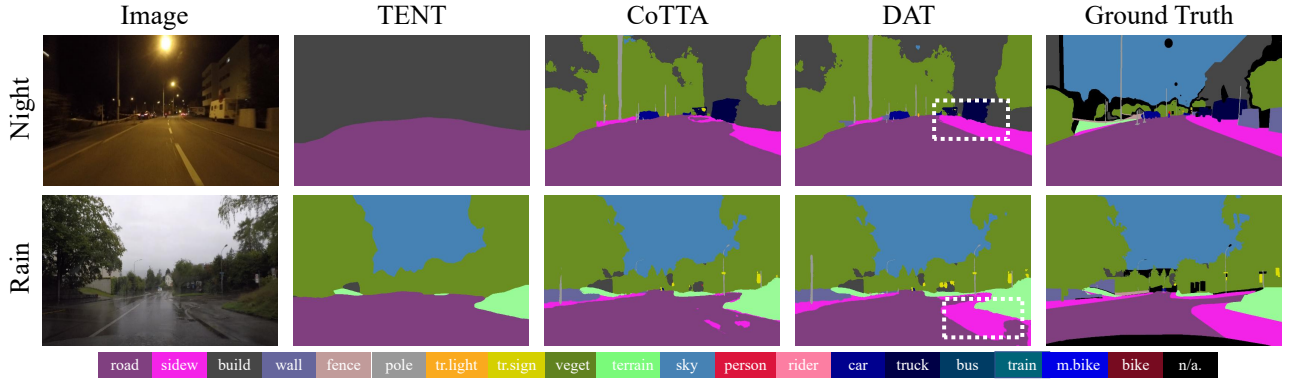
Fig. 3. Qualitative comparisons on the ACDC dataset. Our method could better segment different pixel-wise classes such as shown in the white box.

| | road | sidew | build | wall | fence | pole | tr.light | tr.sign | veget | terrain | sky | person | rider | car | truck | bus | train | m.bike | bike | n/a. |

TABLE III
ABLATION: CONTRIBUTION OF EACH COMPONENT.

	TS	DSP	TRP	mIoU↑	Gain
Ex_1				56.7	/
Ex_2	✓			58.7	+2.0
Ex_3	✓	✓		60.2	+3.5
Ex_4	✓		✓	60.3	+3.6
Ex_5	✓	✓	✓	61.1	+4.4

TABLE IV
EFFECT OF SELECT PARAMETERS' SELECTION METHOD. W/O PAU
MEANS THE SELECTION OF 10% PARAMETERS FROM THE FIRST FRAME IN
THE TARGET SEQUENCE. W PAU DENOTES THE UTILIZATION OF OUR PAU
METHOD TO SELECT AROUND 10% PARAMETERS.

	w/o PAU		w PAU	
Method	mIoU	ACC	mIoU	ACC
Confidence	59.1	71.1	60.6	71.3
Uncertainty	60.3	71.4	61.1	72.2

increases by 1.5%. The outcome underscores the efficacy that fine-tuning domain-specific parameters can effectively mitigate the problem of error accumulation. As shown in Ex_4, TRP achieves further 1.6% mIoU improvement since the task-relevant parameters (TRP) are allocated to positions that are responsive to outputs with minor distribution shifts, which are fine-tuned to avoid the catastrophic forgetting problem. Ex_6 demonstrates the comprehensive integration of all components, resulting in a total mIoU improvement of 4.4%. This underscores the effectiveness of our proposed method and illustrates that all components can collaboratively address the semantic segmentation CTTA problem in autonomous driving.

Percentage of model updated parameters. We conduct an analysis of the impact of the percentage of parameters (DSP and TRP) updated by the DAT method on the CTTA scenario. The percentage of the updated parameters is controlled by the selected parameters (i.e., 0.1%, 0.5%, 1.0%) for each sample. As shown in Figure 4, we gradually increase the percentage of parameters and record the corresponding mIoU values. We find that the mIoU initially improved with the increasing percentage of parameters, but a decline set in when the percentage surpassed 10%. This observation suggests that when the updated parameters are too few, they fail to capture the domain-specific knowledge effectively due to the limited number of parameters. Conversely, if there are too many updated parameters, the model undergoes excessive parameter changes, resulting in catastrophic forgetting. Therefore, it is crucial to strike a balance on the percentage of update parameters, and we consider that DAT can achieve optimal potential in 10% percentage parameters.

Method of select parameters. As shown in Table IV, compared to using the uncertainty method for parameter selection, select parameters based on confidence, the mIoU and ACC
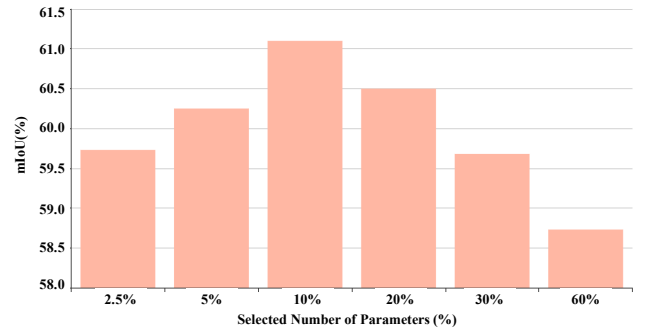

Fig. 4. Different percentages of updated parameters (DSP + TRP)

are significantly lower when using or not using PAU. This further demonstrates that uncertainty can effectively measure domain shift and, consequently, guide parameter selection. Additionally, we observe that in comparison to not employing the PAU strategy, utilizing PAU can further enhance the stability of model updates, subsequently improving model performance. When using PAU and selecting parameters based on uncertainty, the model achieves mIoU and mAcc values of 61.1 and 72.2, respectively.

V. CONCLUSION

Our Distribution-Aware Tuning (DAT) method presents an efficient and practical solution for Semantic Segmentation Continual Test-Time Adaptation (CTTA) in real-world scenarios. DAT intelligently selects and fine-tunes two distinct sets of trainable parameters: domain-specific parameters (DSP) and task-relevant parameters (TRP). This helps mitigate issues of error accumulation and catastrophic forgetting during the continual adaptation process. Given that CTTA in the context of autonomous driving operates on temporal sequences, we introduce a Parameter Accumulation Update (PAU) strategy. This strategy enhances the stability of parameter selection and updates, ensuring a more reliable adaptation process.

REFERENCES

[1] M. Siam, M. Gamal, M. Abdel-Razek, S. Yogamani, M. Jagersand, and H. Zhang, "A comparative study of real-time semantic segmentation for autonomous driving," in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 2018, pp. 587–597.

[2] D. Feng, C. Haase-Schütz, L. Rosenbaum, H. Hertlein, C. Glaeser, F. Timm, W. Wiesbeck, and K. Dietmayer, "Deep multi-modal object detection and semantic segmentation for autonomous driving: Datasets, methods, and challenges," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 3, pp. 1341–1360, 2020.

[3] T. Sun, M. Segu, J. Postels, Y. Wang, L. Van Gool, B. Schiele, F. Tombari, and F. Yu, "Shift: a synthetic driving dataset for continuous multi-task domain adaptation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 21 371–21 382.

[4] C. Sakaridis, D. Dai, and L. Van Gool, "Acdc: The adverse conditions dataset with correspondences for semantic driving scene understanding," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 10 765–10 775.

[5] Q. Wang, O. Fink, L. V. Gool, and D. Dai, "Continual test-time domain adaptation," *ArXiv*, vol. abs/2203.13591, 2022.

[6] J. Liang, R. He, and T. Tan, "A comprehensive survey on test-time adaptation under distribution shifts," *arXiv preprint arXiv:2303.15361*, 2023.

[7] J. Song, J. Lee, I. S. Kweon, and S. Choi, "Ecotta: Memory-efficient continual test-time adaptation via self-distilled regularization," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 11 920–11 929.

[8] Y. Gan, X. Ma, Y. Lou, Y. Bai, R. Zhang, N. Shi, and L. Luo, "Decorate the newcomers: Visual domain prompt for continual test time adaptation," *arXiv preprint arXiv:2212.04145*, 2022.

[9] J. Liu, S. Yang, P. Jia, M. Lu, Y. Guo, W. Xue, and S. Zhang, "Vida: Homeostatic visual domain adapter for continual test time adaptation," *arXiv preprint arXiv:2306.04344*, 2023.

[10] M. Jia, L. Tang, B.-C. Chen, C. Cardie, S. Belongie, B. Hariharan, and S.-N. Lim, "Visual prompt tuning," in *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XXXIII*. Springer, 2022, pp. 709–727.

[11] S. Chen, C. Ge, Z. Tong, J. Wang, Y. Song, J. Wang, and P. Luo, "Adaptformer: Adapting vision transformers for scalable visual recognition," *Advances in Neural Information Processing Systems*, vol. 35, pp. 16 664–16 678, 2022.

[12] Y. Gan, M. Pan, R. Zhang, Z. Ling, L. Zhao, J. Liu, and S. Zhang, "Cloud-device collaborative adaptation to continual changing environments in the real-world," *arXiv preprint arXiv:2212.00972*, 2022.

[13] S. Roy, M. Trapp, A. Pilzer, J. Kannala, N. Sebe, E. Ricci, and A. Solin, "Uncertainty-guided source-free domain adaptation," in *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XXV*. Springer, 2022, pp. 537–555.

[14] Y. Ovadia, E. Fertig, J. Ren, Z. Nado, D. Sculley, S. Nowozin, J. Dillon, B. Lakshminarayanan, and J. Snoek, "Can you trust your model's uncertainty? evaluating predictive uncertainty under dataset shift," *Advances in neural information processing systems*, vol. 32, 2019.

[15] J. Liang, D. Hu, and J. Feng, "Do we really need to access the source data? source hypothesis transfer for unsupervised domain adaptation," in *ICML*, 2020.

[16] D. Chen, D. Wang, T. Darrell, and S. Ebrahimi, "Contrastive test-time adaptation," *ArXiv*, vol. abs/2204.10377, 2022.

[17] M. Boudiaf, R. Mueller, I. B. Ayed, and L. Bertinetto, "Parameter-free online test-time adaptation," *ArXiv*, vol. abs/2201.05718, 2022.

[18] D. Wang, E. Shelhamer, S. Liu, B. A. Olshausen, and T. Darrell, "Tent: Fully test-time adaptation by entropy minimization," in *ICLR*, 2021.

[19] S. Niu, J. Wu, Y. Zhang, Z. Wen, Y. Chen, P. Zhao, and M. Tan, "Towards stable test-time adaptation in dynamic wild world," *arXiv preprint arXiv:2302.12400*, 2023.

[20] L. Yuan, B. Xie, and S. Li, "Robust test-time adaptation in dynamic scenarios," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 15 922–15 932.

[21] I. Shin, Y.-H. Tsai, B. Zhuang, S. Schulter, B. Liu, S. Garg, I. S. Kweon, and K.-J. Yoon, "Mm-tta: Multi-modal test-time adaptation for 3d semantic segmentation," 2022.

[22] J. Song, K. Park, I. Shin, S. Woo, and I. S. Kweon, "Cd-tta: Compound domain test-time adaptation for semantic segmentation."

[23] M. Döbler, R. A. Marsden, and B. Yang, "Robust mean teacher for continual and gradual test-time adaptation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 7704–7714.

[24] Z. Li, S. Shi, B. Schiele, and D. Dai, "Test-time domain adaptation for monocular depth estimation," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2023, pp. 4873–4879.

[25] S. Yang, J. Wu, J. Liu, X. Li, Q. Zhang, M. Pan, and S. Zhang, "Exploring sparse visual prompt for cross-domain semantic segmentation," *arXiv preprint arXiv:2303.09792*, 2023.

[26] A. Tarvainen and H. Valpola, "Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results," *Advances in neural information processing systems*, vol. 30, 2017.

[27] Y. Gal and Z. Ghahramani, "Dropout as a bayesian approximation: Representing model uncertainty in deep learning," in *international conference on machine learning*. PMLR, 2016, pp. 1050–1059.

[28] A. Tarvainen and H. Valpola, "Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results." *Learning*, 2017.

[29] E. Xie, W. Wang, Z. Yu, A. Anandkumar, J. M. Alvarez, and P. Luo, "Segformer: Simple and efficient design for semantic segmentation with transformers," *Advances in Neural Information Processing Systems*, vol. 34, pp. 12 077–12 090, 2021.

[30] D. Wang, E. Shelhamer, S. Liu, B. Olshausen, and T. Darrell, "Tent: Fully test-time adaptation by entropy minimization," *arXiv preprint arXiv:2006.10726 , year=2020*.

[31] Y. Gao, X. Shi, Y. Zhu, H. Wang, Z. Tang, X. Zhou, M. Li, and D. N. Metaxas, "Visual prompt tuning for test-time domain adaptation," *arXiv preprint arXiv:2210.04831*, 2022.

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

[33] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[34] J. Liu, Q. Zhang, J. Li, M. Lu, T. Huang, and S. Zhang, "Unsupervised spike depth estimation via cross-modality cross-domain knowledge transfer," *arXiv preprint arXiv:2208.12527*, 2022.