

OccTTA: Fast and Memory-Efficient Test-Time Adaptation for 3D Occupancy Prediction

Peixin Chen, Wei Zhang, Zijian He, Yipeng Qin, Qiqing Lao, Jie Feng, Guanbin Li *Member, IEEE*

Abstract—3D semantic occupancy prediction (OCC) is a crucial task in vision-centric autonomous driving, focusing on reconstructing fine-grained 3D geometry and semantics of the scene surrounding the ego vehicle. Despite recent progress, OCC still faces a significant challenge in generalizing to real-world scenarios due to the unpredictable domain shifts between training data and diverse testing environments. A promising solution to this challenge is test-time adaptation (TTA), which enhances model generalization by updating its parameters to align with the testing domain during inference. However, applying existing TTA methods to OCC is non-trivial as: i) Practical deployment of OCC on resource-constrained onboard devices imposes requirements on memory efficiency. ii) OCC must rapidly adapt to dynamic scenes in latency-sensitive autonomous driving scenarios, posing a challenge for fast test-time adaptation. To address the above issues, we propose a novel fast and memory-efficient Occupancy Test-Time Adaptation (OccTTA) method. Specifically, i) to achieve memory-efficient TTA, we propose a novel TTA framework for OCC based on lite-boost units; ii) to enable fast TTA, we propose a novel forward-backward adaptation strategy that synergizes computationally lightweight forward updates with selective backward optimization. Specifically, our strategy selectively triggers backward updates only for divergent inputs, while employing a k-NN classifier to perform efficient forward updates for accelerating the model's responsiveness to domain shifts, thereby simultaneously improving adaptation efficiency and achieving superior performance. Moreover, to address the class imbalance problem in OCC, we introduce an adaptive class-balanced loss that dynamically prioritizes challenging classes, enabling more robust TTA. Extensive experiments on the nuScenes-C benchmark verify the effectiveness of our method.

Index Terms—Test-time Adaptation, 3D Occupancy Prediction, Autonomous Driving.

I. INTRODUCTION

Vision-based 3D semantic occupancy prediction (OCC) aims to estimate the occupancy states and semantic labels of 3D voxels surrounding the ego vehicle for a comprehensive understanding of 3D scenes. Compared to traditional perception tasks, such as 3D object detection [1]–[5] and bird's-eye view

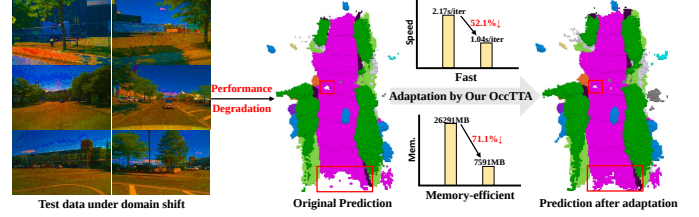


Fig. 1: OccTTA mitigates domain-shift-induced performance degradation in 3D occupancy prediction in a fast and memory-efficient manner.

(BEV) semantic segmentation [6]–[9], it offers finer-grained geometric perception and general object recognition capabilities, eliciting significant interest among the autonomous driving community. Previous works [10]–[22] have extensively investigated various aspects of this task, including the accuracy, efficiency, and memory overhead, achieving superior performance.

Despite significant progress, existing OCC methods rely on an ideal assumption: *test images come from the same distribution as training images*. However, this assumption does not hold true in real-world scenarios, where autonomous vehicles encounter natural variations such as weather-induced lighting changes or sensor noise, causing domain shifts between training and test data. These shifts cause significant performance degradation in well-trained models, leading to unintended traffic accidents and severe safety risks, which limit their practical utility (Fig. 2). In this work, we address the above-mentioned gap by proposing **Occupancy Test-Time Adaptation (OccTTA)**, the first approach to apply Test-Time Adaptation (TTA) to the OCC task. However, this application is non-trivial. While TTA allows models to adapt to unlabeled target domain data, existing methods [23]–[37] primarily address general domain shifts and overlook two critical challenges specific to OCC: i) Practical deployment of OCC on resource-constrained onboard devices imposes requirements on **memory efficiency**. ii) OCC models demand rapid adaptation to dynamic environmental shifts under latency-sensitive deployment settings, posing a challenge for **fast** test-time adaptation. Therefore, conventional TTA methods that primarily adjust parameters of normalization layers (*e.g.*, BatchNorm) are not applicable as: i) Updating normalization parameters in deep networks introduces significant memory overhead, as it requires storing large intermediate activations, conflicting with the need for memory efficiency. ii) Conventional TTA methods, which rely on iterative back-propagation per sample, incur prohibitive computational overhead, hindering their deployment in latency-sensitive autonomous driving

• Peixin Chen, Zijian He, Qiqing Lao and Guanbin Li are with the School of Computer Science and Engineering, Sun Yat-sen University, Guangzhou 510006, China. Peixin Chen and Guanbin Li are also with Peng Cheng Laboratory, Shenzhen 518066, China and also with Guangdong Key Laboratory of Big Data Analysis and Processing, Guangzhou, 510006, China. E-mail: chenpx28@mail2.sysu.edu.cn, hezj39@mail2.sysu.edu.cn, laoqq@mail2.sysu.edu.cn, liguanbin@mail.sysu.edu.cn.

• Wei Zhang is with the Baidu Inc., China. E-mail: zhangwei99@baidu.com.

• Yipeng Qin is with the School of Computer Science and Informatics, Cardiff University, Cardiff, UK. Email: qiny16@cardiff.ac.uk.

• Jie Feng is with the Key Laboratory of Intelligent Perception and Image Understanding of Ministry of Education of China, Xidian University, Xi'an 710071, China. Email: jiefeng0109@163.com

• Corresponding author: Guanbin Li.

scenarios. Addressing these challenges, our OccTTA proposes two novel techniques (Fig. 1): i) To achieve memory-efficient TTA, we propose a novel efficient TTA framework for OCC based on lite-boost units. By integrating lite-boost units into the OCC model, our framework minimizes memory consumption through reduced intermediate activation sizes. ii) To enable fast TTA, we propose a novel forward-backward adaptation strategy that significantly reduces adaptation latency while simultaneously enhancing model performance. Our strategy selectively triggers computationally intensive backward updates only for divergent inputs, while employing a k-NN classifier to perform efficient forward updates for accelerating the model’s responsiveness to domain shifts, thereby simultaneously improving adaptation efficiency and achieving superior performance. In addition to the aforementioned challenges, OCC models face severe class imbalance in the pseudo-labels generated during adaptation. Directly adapting to these imbalanced pseudo-labels risks pushing the model toward trivial solutions, hindering adaptation effectiveness. To address this issue, we introduce an adaptive class-balanced loss that dynamically prioritizes adaptation towards more challenging classes based on online difficulty measurements, ensuring stable adaptation to imbalanced class distributions.

Our main contributions can be summarized as follows:

- We present **OccTTA**, the first test-time adaptation approach tailored for OCC models, boosting their generalization to real-world scenarios.
- To ensure OccTTA meets the memory efficiency requirement of OCC, we propose a novel framework that leverages lightweight lite-boost units to enable residual learning, thereby minimizing intermediate activation sizes.
- To enhance the efficiency of OccTTA, we propose a novel forward-backward adaptation strategy that synergizes computationally lightweight forward updates with selective computationally intensive backward optimization. We also propose a novel adaptive class-balanced loss to address OCC’s class imbalance problem.
- Extensive experiments on the nuScenes-C benchmark demonstrate the superiority of OccTTA, achieving 23.0% higher mIoU over the state-of-the-art TTA method under severe domain shifts, along with a 71.1% reduction in memory overhead and a 52.1% acceleration in adaptation speed.

II. RELATED WORK

A. 3D Semantic Occupancy Prediction

3D semantic occupancy prediction (OCC) task has garnered significant attention in recent years due to its enhanced geometry information and superior capability in detecting general objects. TPVFormer [11] projects surrounding multi-camera features into a tri-perspective view representation and leverages sparse point cloud supervision for 3D occupancy prediction. SurroundOcc [10] employs a multi-scale transformer architecture to construct 3D voxel representations and hierarchically fuses them in a coarse-to-fine manner. COTR [13] proposes a compact voxel representation to reduce computational costs. It

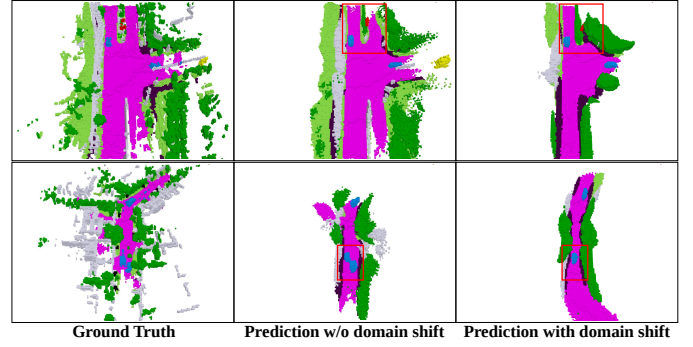


Fig. 2: **Performance degradation:** Domain shifts between training data and testing environments result in significant performance degradation that manifests as structural distortions (e.g., narrowed lanes, merged vehicles, highlighted in red boxes).

employs an explicit-implicit view transformation mechanism to generate geometry-aware features and adopts a coarse-to-fine semantic grouping strategy to enhance semantic discriminability. PanoOcc [17] introduces a unified occupancy representation for 3D panoptic segmentation via a coarse-to-fine learning scheme. GaussianFormer [22] models 3D scenes with sparse 3D semantic gaussians and designs a gaussian-to-voxel splatting module to generate the semantic occupancy. OccNeRF [21] integrates neural radiance fields with occupancy prediction, leveraging multi-view geometric constraints to reconstruct dense 3D occupancy fields. SparseOcc [15] introduces a fully sparse occupancy network that reconstructs a sparse 3D representation and predicts occupancy through sparse queries, effectively reducing computational overhead. FastOcc [14] accelerates the 3D occupancy prediction process by simplifying 3D convolution blocks to a 2D BEV convolution network and completing the BEV features with the interpolated voxel features. STCOcc [38] introduces a sparse spatial-temporal cascade renovation framework that leverages predicted occupied states to guide 3D feature refinement and sparse temporal modeling, reducing memory and computational costs for occupancy prediction. S2GO [39] represents driving scenes with a compact set of streaming sparse 3D queries and decodes them into semantic Gaussians, achieving efficient occupancy prediction with faster inference.

While the aforementioned methods have made significant progress in occupancy prediction, they remain constrained by the assumption of identical training-test data distributions, overlooking the critical challenge of domain shifts in real-world deployment. Our method presents an early effort to address domain shift challenges for OCC models, establishing a foundation for robust cross-domain generalization.

B. Test-Time Adaptation

Test-Time Adaptation (TTA) aims to improve model performance on test data via model adaptation through unlabeled test samples. TENT [23] introduces a fully test-time adaptation method that updates batch normalization parameters via entropy minimization. CoTTA [25] proposes continual TTA setup and applies weight-averaged and augmentation-averaged predictions with stochastic weight restoration to mitigate error

accumulation and catastrophic forgetting. MEMO [26] minimizes the marginal entropy of predictions across different augmented views of a given test sample to encourage the model to make consistent predictions across different augmentations. SAR [28] investigates the effects of norm layers in TTA and introduces a sharpness-aware and reliable entropy minimization method to achieve stable model adaptation. RoTTA [27] utilizes the teacher-student model to stabilize adaptation and selects the data with category-balanced sampling with timelessness and uncertainty. COME [40] introduces conservative entropy minimization for robust adaptation, which models prediction uncertainty with a Dirichlet prior and regularizes the model to favor conservative confidence on unreliable samples. ReCAP [41] models region-level uncertainty to construct a proxy objective for region confidence optimization, effectively mitigating noisy adaptation dynamics and achieving superior generalization. [35] introduces hierarchical distribution alignment, performing style-level correction in shallow layers and semantic-level alignment in deeper layers, achieving fast and stable test-time generalization. A3-TTA [34] proposes an adaptive anchor alignment framework that constructs reliable pseudo-labels through anchor-guided supervision, effectively mitigating pseudo-label noise, feature misalignment, and error accumulation in TTA. COLA [36] proposes a context-aware, language-driven TTA approach that leverages semantic textual information to guide feature alignment, improving robustness under distribution shifts. [37] proposes an uncertainty- and distribution-aware reset strategy, demonstrating that properly scheduling and locating resets is crucial for suppressing error accumulation and ensuring long-term adaptation stability. Conventional TTA methods mainly concentrate on image

classification and segmentation tasks, and recent efforts have extended TTA to other domains. MonoTTA [33] adapts TTA to monocular 3D object detection by designing an adaptive strategy to identify reliable objects for model adaptation and introducing a negative regularization term to leverage low-score objects through negative learning. BATCLIP [42] enhances CLIP’s robustness to common image corruptions by jointly adapting its visual and text encoders within the shared multimodal feature space. BCA [43] continuously updates text-based class embeddings to align likelihoods with incoming image features and concurrently refines class priors using posterior estimation. FGA [44] introduces flatness-aware optimization into TTA, guiding models toward flatter loss landscapes to enhance generalization and robustness of vision-language models under distribution shifts.

However, applying existing TTA methods to OCC is non-trivial as: i) Practical deployment of OCC on resource-constrained onboard devices imposes requirements on memory efficiency. ii) OCC must rapidly adapt to dynamic scenes under latency-sensitive deployment settings, posing a challenge for fast test-time adaptation. To the best of our knowledge, our OccTTA is the first to extend TTA to the OCC task and address these challenges, enabling fast and memory-efficient TTA in real-world scenarios.

III. PRELIMINARIES

TTA Problem Formulation. Given a pre-trained model $f_{\Theta}(\mathbf{x})$ with parameters Θ , trained on labeled training images $\{(\mathbf{x}_i^s, \mathbf{y}_i^s)\}_{i=1}^N$, where N is the number of training samples and $\mathbf{x}^s \sim P(\mathbf{x})$ follows the training distribution $P(\mathbf{x})$. During training, the model is optimized to fit (or overfit)

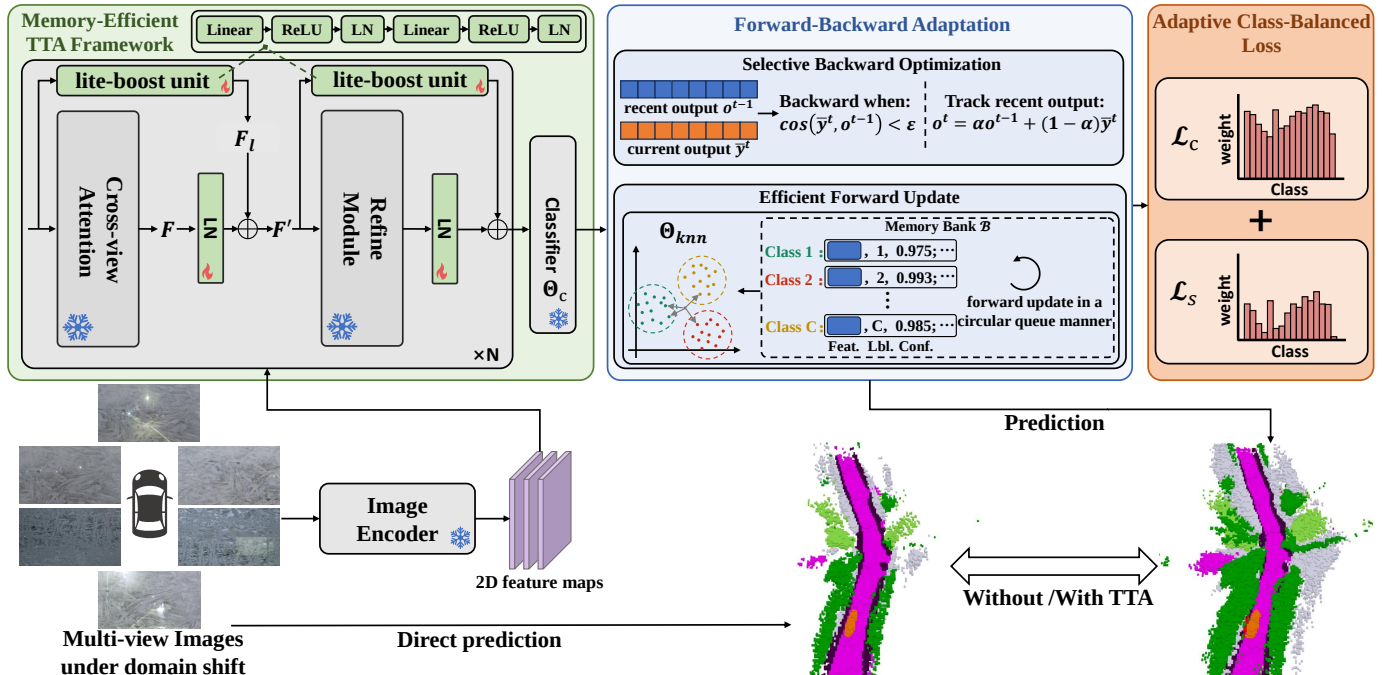


Fig. 3: **The pipeline of our OccTTA.** Our OccTTA comprises three components: i) A memory-efficient TTA framework based on lite-boost units (green). ii) A forward-backward adaptation strategy that combines efficient forward updates and selective backward optimization (blue). iii) An adaptive class-balanced loss (orange). The original OCC model is represented in gray.

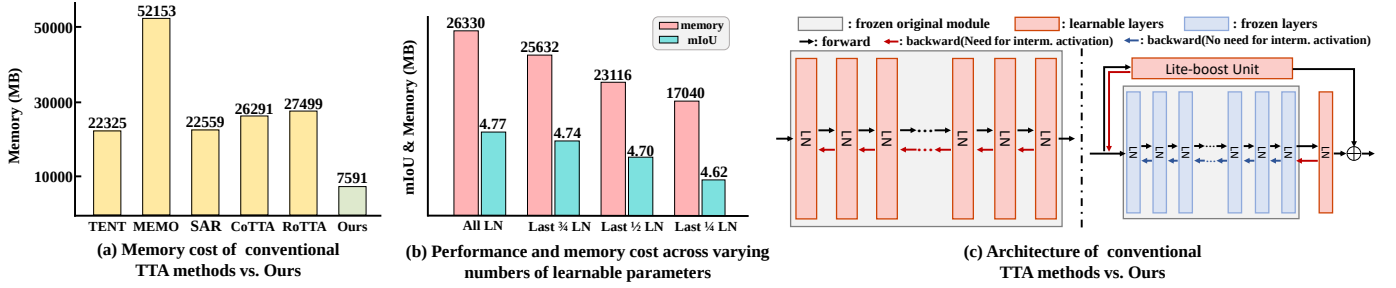


Fig. 4: Analysis of conventional TTA methods for OCC Models.

the training data. However, during testing, the performance of $f_{\Theta}(\mathbf{x})$ would heavily degrade once there are distribution shifts between training and test images, *i.e.*, testing images $\mathbf{x} \sim Q(\mathbf{x})$ and $P(\mathbf{x}) \neq Q(\mathbf{x})$. Our goal is to enhance the model's performance during test time through continuous target domain adaptation in an online manner. Specifically, we aim to update the model by minimizing unsupervised objectives defined on test samples:

$$\min_{\hat{\Theta}} \mathcal{L}(p) \quad (1)$$

where $\mathcal{L}$ is the loss function, $\hat{\Theta} \in \Theta$ denotes the subset of model parameters that should be updated, p is the prediction on the test images $\mathbf{x}$.

IV. METHOD

A. Overview

To achieve TTA for OCC, we propose a novel Occupancy Test-Time Adaptation (OccTTA) method, which includes: i) A memory-efficient TTA framework based on lite-boost units that minimizes memory overhead by reducing intermediate activation storage. ii) A forward-backward adaptation strategy that enhances adaptation efficiency by reducing redundant and time-consuming back-propagation steps, while incorporating efficient forward updates to accelerate the model's responsiveness to domain shifts. iii) An adaptive class-balanced loss that dynamically prioritizes challenging classes to address class imbalance in OCC, further improving TTA performance. Please see Fig. 3 for an overview of the pipeline of our OccTTA method.

B. Memory-Efficient TTA Framework

As mentioned above, test-time adaptation (TTA) for OCC models is inherently constrained by limited memory resources, rendering conventional TTA methods that primarily update normalization layers impractical due to their substantial memory overhead (Fig. 4(a)). A straightforward workaround is to reduce the number of learnable parameters by only adapting the deep normalization layers (*e.g.*, LayerNorm, LN) in the model. However, this results in performance degradation and offers only marginal memory savings (Fig. 4(b)). This limitation arises because memory consumption during training is dominated by *intermediate activations* rather than learnable parameters [45].

Memory Overhead Analysis of LayerNorm. Theoretically, let $\mathbf{a}_i$ and $\mathbf{b}_i$ denote the scale and shift factors of the i -th

LayerNorm layer of the OCC model, $\mathbf{F}_i$ its normalized input feature, and $\mathbf{F}_{i+1}$ its output feature. The forward computation is defined as:

$$\mathbf{F}_{i+1} = \mathbf{a}_i \mathbf{F}_i + \mathbf{b}_i \quad (2)$$

with backward gradients computed as:

$$\frac{\partial \mathcal{L}}{\partial \mathbf{F}_i} = \frac{\partial \mathcal{L}}{\partial \mathbf{F}_{i+1}} \mathbf{a}_i, \quad \frac{\partial \mathcal{L}}{\partial \mathbf{a}_i} = \frac{\partial \mathcal{L}}{\partial \mathbf{F}_{i+1}} \mathbf{F}_i, \quad \frac{\partial \mathcal{L}}{\partial \mathbf{b}_i} = \frac{\partial \mathcal{L}}{\partial \mathbf{F}_{i+1}} \quad (3)$$

Eq. 3 demonstrates that updating each LN layer requires storing intermediate activations $\mathbf{F}_i$. Given a large OCC model, conventional TTA methods incur substantial memory overhead due to the storage of large intermediate activations for multiple LN layers. To address this, we propose a memory-efficient TTA framework that minimizes memory consumption via lightweight residual update (Fig. 4(c)).

Memory-Efficient Residual Update. Specifically, we attach lite-boost units and auxiliary LN layers *external* to the original cross-view attention modules and refine modules of the OCC model, thereby enabling lightweight residual update. Let $\mathbf{F}_l$ be the output of the lite-boost unit, and $\mathbf{F}$ the output of the cross-view attention module (or refine module), we have the refined feature $\mathbf{F}'$ as follows:

$$\mathbf{F}' = \text{LN}(\mathbf{F}) + \mathbf{F}_l \quad (4)$$

This design offers dual advantages: i) the lite-boost units minimize memory overhead by reducing the size of intermediate activations required for back-propagation, enabling memory-efficient adaptation; ii) freezing original model parameters ensures robust adaptation without sacrificing pre-trained knowledge, avoiding catastrophic forgetting. To ensure robust adaptation, we employ a teacher-student framework similar to CoTTA [25].

Remark. Note that the key contribution of our framework lies not in the specific internal design of the lite-boost units, but in the architectural strategy of externally attaching these units to the modules of the original model. Therefore, without loss of generality, we implement each lite-boost unit as a lightweight module comprising two Linear-ReLU-LN blocks. Please see Table VII for more details.

C. Forward-Backward Adaptation Strategy

OCC models demand rapid adaptation to dynamic domain shifts in latency-sensitive autonomous driving scenarios, yet existing methods suffer from high computational costs due to per-sample iterative backward optimization, posing a challenge for fast test-time adaptation. To address this challenge, we

propose a novel forward-backward adaptation strategy that jointly leverages computationally efficient forward and backward pathways to optimize predictions with minimal latency.

1) *Selective Backward Optimization*: For backward adaptation, we exploit a key observation in autonomous driving: test data evolves gradually, and adjacent frames often yield similar gradients. As shown in Fig. 6, our empirical statistics reveal that over 50% of consecutive frames exhibit gradient similarities exceeding 99%. To enhance adaptation efficiency, we selectively skip the backward loss and gradient computations for samples whose output cosine similarity with recent predictions exceeds a predefined threshold ϵ :

$$S(\bar{y}^t) = \mathbb{I}_{\{\cos(\bar{y}^t, o^{t-1}) < \epsilon\}}(\bar{y}^t) \quad (5)$$

$$o^t = \alpha o^{t-1} + (1 - \alpha) \bar{y}^t \quad (6)$$

where $\bar{y}^t = \frac{1}{N'} \sum_{i=1}^{N'} p_i^t$ represents the average prediction over N' non-empty voxels at time step t , and p_i^t denotes the model's prediction for voxel i at step t ; o^t is the moving average of model outputs, dynamically tracking recent predictions. In this way, we ensure that backward optimization is executed only when $S(\bar{y}) = 1$ (i.e., when the cosine similarity falls below ϵ), thereby removing redundant computations while preserving adaptation efficacy.

2) *Efficient Forward Update*: On the other hand, while the teacher-student framework enables robust adaptation, its inherent reliance on gradual model updates results in its slow response to abrupt domain shifts—an issue exacerbated when combined with selective backward optimization. To address this, unlike conventional methods that solely rely on backward optimization, we simultaneously enhance predictions via lightweight forward updates, enabling fast responsiveness to domain shifts while achieving superior performance. Specifically, we additionally integrate a k-NN classifier Θ_{knn} that directly refines predictions during each forward pass. Θ_{knn} is a non-parametric module that maintains a memory bank $\mathcal{B}$ of representative features f , labels l , and confidence scores c , and makes predictions at inference time via *confidence-weighted similarity voting* by retrieving the top- k nearest neighbors and aggregating their predictions as follows:

$$\Theta_{knn}(f_i) = \text{softmax}\left(\sum_{j \in \mathcal{B}_k} \mu_{ij} \cdot c_j \cdot l_j\right) \quad (7)$$

where μ_{ij} denotes the cosine similarity between the input feature f_i and memory feature f_j , c_j is the confidence score associated with f_j , l_j is its one-hot encoded class label, and $\mathcal{B}_k$ denotes the top- k nearest neighbors retrieved from the memory bank. After that, we refine the final prediction p_i by combining the output of the original OCC classifier Θ_c with that of Θ_{knn} :

$$p_i = \lambda \Theta_c(f_i) + (1 - \lambda) \Theta_{knn}(f_i) \quad (8)$$

where $\lambda \in [0, 1]$ is a balancing coefficient.

Memory Bank $\mathcal{B}$ Update Protocol. The k-NN classifier rapidly adapts to domain shifts through immediate feature memory updates, delivering performance gains during early adaptation phases. Specifically, we introduce a memory bank update protocol. Each class is assigned an equal number of memory slots to ensure class balance, and during online

updates, for each sample's prediction, the top- k highest-confidence features per class are selected to update their respective memory queues in a circular queue manner. This approach enables the memory bank to evolve with the test-time distribution while incurring minimal computational overhead. Furthermore, to preserve critical source domain knowledge, we partially freeze the memory bank by retaining a subset of source domain representations, thereby striking a balance between adaptability and knowledge retention during target domain adaptation. Please see Fig. 5 for a clear and intuitive understanding of our memory update protocol.

Remark. The integration of Θ_c and Θ_{knn} enables the model to accelerate its response to domain shifts while preserving adaptation stability. By synergizing selective backward optimization with computationally efficient forward updates, our method concurrently achieves superior adaptation speed and prediction accuracy.

D. Adaptive Class-Balanced Loss

Conventional TTA methods often assume that the classwise data distributions are well balanced, rendering them ill-suited for real-world OCC scenarios, where confidence levels across classes vary significantly (Fig. 7(a)) and severe class imbalance (e.g., >90% empty-class voxels, Fig. 7(b)) often triggers trivial solution convergence. To achieve robust adaptation under class-imbalanced conditions, we design an adaptive class-balanced loss incorporating a dynamic difficulty-aware reweighting mechanism. Firstly, we compute class-wise confidence levels by temporally accumulating per-class prediction confidence. For class c at time step t , we have:

$$\bar{p}^t[c] = \frac{\sum_{i=1}^N \mathbb{I}_{\{p[c] > \text{top}2(p)\}}(p_i^t) \cdot p_i^t[c]}{\sum_{i=1}^N \mathbb{I}_{\{p[c] > \text{top}2(p)\}}(p_i^t)} \quad (9)$$

$$m^t[c] = \beta m^{t-1}[c] + (1 - \beta) \bar{p}^t[c] \quad (10)$$

where $p_i^t[c]$ denotes the predicted probability of class c for voxel v_i at step t ; $\mathbb{I}_{\{p[c] > \text{top}2(p)\}}$ is an indicator function that selects the first two high confidence predictions to increase the tolerance against noise in pseudo labels, which is inspired by [46]; $\bar{p}^t[c]$ denotes the averaged probability value of class c at step t ; $m^t[c]$ denotes the confidence level of class c at step t .

Then, we define the adaptive class-balanced loss leveraging the following insight: confidence levels indicate the relative difficulty of different classes, with lower confidence corresponding to inherently more challenging classes that require increased model focus. Accordingly, we have:

$$\mathcal{L}_c(q_i^t) = - \sum_{c=1}^C (1 - \frac{1}{2}(\bar{p}^t[c])^2) p_i^t[c] \log q_i^t[c] \quad (11)$$

$$\mathcal{L}_s(q_i^t) = - \sum_{c=1}^C (1 - \bar{p}^t[c]) \log q_i^t[c] \quad (12)$$

where p_i^t and q_i^t denote the predictions of the teacher model and student model for voxel i at step t , respectively; $\mathcal{L}_c$ is a weighted cross-entropy loss that assigns higher weights

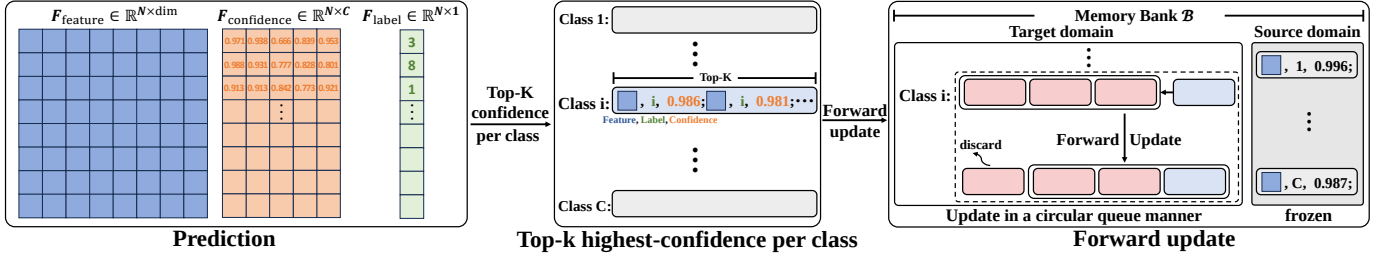


Fig. 5: Illustration of memory bank update protocol (take class i for example).

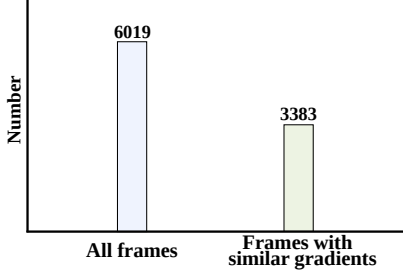


Fig. 6: Statistics on the number of frames with similar gradients. Experiments are conducted on the nuScenes-C validation set with Gaussian noise corruption.

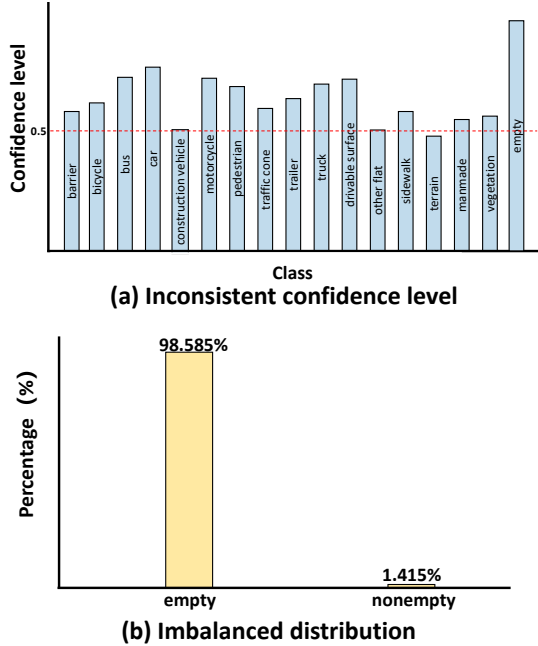


Fig. 7: (a) **Inconsistent confidence level:** significant variations in confidence levels exist across classes. (b) **Imbalanced distribution:** the predictions of OCC model exhibit severe class imbalance due to geometry sparsity, with over 90% of predicted voxels belonging to the empty class.

to under-learned classes while suppressing overconfident predictions for dominant classes; $\mathcal{L}_s$ is a confidence-guided suppression mechanism inspired by [47]: predictions for high-confidence classes are suppressed, which redistributes learning focus to other low-confidence classes. Our adaptive class-balanced loss ensures stable adaptation to imbalanced distributions by dynamically reweighting both supervision signals and prediction constraints.

Overall Loss. To mitigate error accumulation and catastrophic forgetting, we incorporate a regularization term $\mathcal{L}_r$ to constrain the refined feature from deviating significantly from the output feature of the frozen original networks:

$$\mathcal{L}_r = |\mathbf{F}' - \mathbf{F}| \quad (13)$$

However, setting the regularization strength presents a dilemma: excessive constraints hinder effective learning, while insufficient constraints fail to suppress error propagation. To resolve this, we set a small regularization constraint and additionally record a less inert moving average n^t of confidence levels to estimate dynamic shifts in sample distributions. Specifically, we compute:

$$n^t = \gamma n^{t-1} + (1 - \gamma) \bar{p}^t \quad (14)$$

where n^t represents the distributional shift estimate at step t . With $\gamma < \beta$, n^t exhibits a higher sensitivity to distributional changes compared to m^t . We define a large threshold ϵ' ; when $\sum_{c=1}^C |n^t[c] - m^t[c]| > \epsilon'$, indicating abrupt distribution shifts, we reset the parameters of lite-boost units to prevent error accumulation caused by outdated adaptations. The overall loss function is formulated as below:

$$\mathcal{L} = \mathcal{L}_c + \lambda_s \mathcal{L}_s + \lambda_r \mathcal{L}_r \quad (15)$$

where λ_s and λ_r are trade-off parameters.

V. EXPERIMENT

Experimental setup. Following CoTTA [25], we conduct experiments on the continual TTA task, where the deployed model is continually adapted to each corruption type sequentially. We empirically validate whether the proposed OccTTA method can effectively enhance the performance of OCC models under a batch size of 1. The rationale for this setup is that in real-world autonomous driving scenarios, OCC methods require fast adaptation to dynamically evolving environments to support latency-sensitive online prediction, where *sequential test-time updates must operate on single-frame inputs*. It is noteworthy that this experimental setup poses significant challenges due to constrained mini-batch size, correlated sampling dependency, non-static distribution shifts, and pronounced class imbalance, which collectively exacerbate adaptation instability.

Dataset. nuScenes [48] is a large-scale autonomous driving dataset, which contains 700 training sequences and 150 validation sequences. Each sequence lasts 20 seconds with RGB images collected by 6 surround-view cameras. We utilize the dense semantic occupancy annotations from SurroundOcc [10] for evaluation, where the 3D voxel space spans [-50m,-50m,-3m] to [50m,50m,5m] with a resolution of $0.5m^3$. Each voxel

TABLE I: Comparisons with state-of-the-art methods on nuScenes-C validation set (**SEVERITY LEVEL 5**) with **BATCH SIZE=1** regarding mIoU, where experiments are based on GaussianFormer [22]. **bold**: the best result; underline: second best result. * denotes the implementation incorporating our adaptive class-balanced loss.

Method	Noise			Blur			Weather				Digital			Avg.
	Gauss.	Shot	Impul.	Defoc.	Glass	Motion	Snow	Frost	Fog	Brit.	Contr.	Pixel	Sat.	
Source-only	2.77	4.02	2.89	5.80	17.21	8.95	2.29	8.21	15.88	16.85	5.80	15.83	8.19	8.82
TENT	0.01	-	-	-	-	-	-	-	-	-	-	-	-	fail
MEMO	-	-	-	-	-	-	-	-	-	-	-	-	-	fail
COME	-	-	-	-	-	-	-	-	-	-	-	-	-	fail
SAR	1.36	0.60	0.13	0.16	0.63	0.05	0.00	0.03	0.05	0.04	0.00	0.02	0.00	0.24
CoTTA	2.72	3.91	2.79	5.65	16.17	8.12	2.04	7.28	14.69	15.61	5.17	15.01	7.43	8.20
CoTTA*	<u>3.47</u>	<u>4.99</u>	<u>3.69</u>	<u>6.42</u>	16.54	9.15	<u>3.04</u>	8.54	15.67	16.60	<u>6.38</u>	16.03	<u>8.46</u>	<u>9.15</u>
RoTTA	2.82	4.17	3.11	5.83	13.93	7.98	2.10	5.56	9.19	9.19	3.77	6.77	3.47	5.99
RoTTA*	2.89	4.30	3.22	6.06	<u>17.42</u>	<u>9.23</u>	2.39	<u>8.61</u>	<u>16.07</u>	<u>17.08</u>	6.05	<u>16.17</u>	8.36	9.07
OccTTA(Ours)	3.51	5.57	4.28	7.14	18.46	10.26	3.25	9.50	16.34	18.23	7.51	17.35	9.81	10.09

TABLE II: Comparison of memory overhead and adaptation speed.

Method	TENT	MEMO	COME	SAR	CoTTA	RoTTA	OccTTA(Ours)
Mem.(MB)	22325	52153	22325	22559	26291	27499	7591
Speed(s/iter)	1.31	4.31	1.32	2.63	2.17	2.23	1.04

is labeled across 18 categories, including 16 object classes, an “unknown” class, and an “empty” class. Following [33], we construct the nuScenes-C benchmark through the incorporation of 13 distinct types of data corruptions [49] at 5 severity levels.

Compared Methods. We compare our method with the following methods: 1) source-only directly applies the pre-trained GaussianFormer [22] model to the test data within corruptions without adaptation; 2) Tent [23] minimizes the entropy of test samples; 3) SAR [28] selectively prunes high-entropy samples and optimizes the entropy loss to a flat minimum; 4) COME [40] introduces conservative entropy minimization for robust adaptation; 5) MEMO [26] enforces the prediction consistency of different augmented copies of a test sample; 6) CoTTA [25] employs a mean teacher framework to minimize cross-entropy loss while integrating stochastic partial weight restoration; 7) RoTTA [27] leverages a mean teacher framework to optimize a temporally-weighted cross-entropy loss and selects the data with category-balanced sampling with timeliness and uncertainty.

Implementation Details and Evaluation Metric. OccTTA is a general framework that can be applied to most existing pre-trained OCC models, endowing them with robustness to domain shift. Without loss of generality, we choose GaussianFormer [22] as the source model and conduct model adaptation based on publicly available pre-trained weights provided by its authors. Similar to previous TTA works [29]–[32], we pre-train the lite-boost units on the source dataset before deployment. We use a cross-entropy loss with an Adam optimizer with a learning rate of $2e-4$ to pre-train the lite-boost units for only one epoch. Note that we do not require the source dataset during test time. The k-NN classifier’s memory bank is initialized with 1,000 slots per class and populated with source domain representative features by maximizing intra-class feature space coverage through farthest feature selection relative to class centroids. During adaptation, we employ an Adam optimizer with a learning rate of $2e-4$ and no weight

decay. The memory bank is updated through forward updates with 300 slots per class. The top- K parameter is set to 10 for confidence-weighted similarity voting and 50 for memory bank updates. The moving average factors α , β , and γ in Eq. 6, Eq. 10, and Eq. 14 are set to 0.9, 0.99, and 0.9, respectively. The online confidence level m^t and the less inert one n^t in Eq. 10 and Eq. 14 are initialized to $1/\mathcal{C}$ at $t = 0$, where $\mathcal{C}$ denotes the total number of semantic categories. The skip threshold ϵ and reset threshold ϵ' are set to 0.985 and 2.0, respectively. The trade-off parameter λ in Eq. 8 is assigned a value of 0.01. The trade-off parameters λ_s and λ_r in Eq. 15 are assigned values of 0.0001 and 0.05, respectively. We report our experimental results with the mean intersection over union (mIoU) of all semantic classes.

A. Comparisons with State-of-the-arts

We compare our approach to competing TTA methods on nuScenes-C under the highest severity level 5, with quantitative results detailed in Table I and Table II. Due to distribution shifts, directly applying the pre-trained model (*i.e.*, source-only) suffers severe performance degradation across all corruption types. Competing TTA methods struggle to improve model performance, underscoring the unique challenges of adapting OCC models to domain shifts. In contrast, our method consistently achieves state-of-the-art or competitive performance with superior efficiency. Specifically, our method achieves 23.0% higher mIoU over the best counterpart method CoTTA [25] under severe domain shifts, along with a 71.1% reduction in memory overhead and a 52.1% acceleration in adaptation speed. The original CoTTA and RoTTA frameworks exhibit limited efficacy in enhancing model performance, whereas their variants equipped with our adaptive class-balanced loss achieve better performance, demonstrating the effectiveness of our adaptive class-balanced loss. Other baselines (TENT [23], MEMO [26], SAR [28], COME [40]) trigger model collapse (*i.e.*, predicting only empty classes). We attribute this failure to two factors: (1) The high gradient

TABLE III: Adaptation performance on nuScenes-C validation set using SurroundOcc [10].

Method	Noise			Blur			Weather				Digital			Avg.
	Gauss.	Shot	Impul.	Defoc.	Glass	Motion	Snow	Frost	Fog	Brit.	Contr.	Pixel	Sat.	
Source-only	8.01	9.09	8.48	8.86	18.85	9.19	3.93	7.93	17.19	18.22	5.44	15.29	11.37	10.91
TENT	0.16	-	-	-	-	-	-	-	-	-	-	-	-	fail
SAR	2.56	-	-	-	-	-	-	-	-	-	-	-	-	fail
CoTTA	8.01	9.09	8.48	8.86	18.85	9.19	3.93	7.93	17.19	18.22	5.44	15.29	11.37	10.91
RoTTA	8.01	9.07	8.46	8.89	18.81	9.26	3.83	7.98	17.04	18.03	5.25	14.92	11.08	10.82
OccTTA(Ours)	8.86	10.08	9.43	9.57	19.98	10.11	4.59	8.95	18.82	19.56	5.96	17.35	13.53	12.06

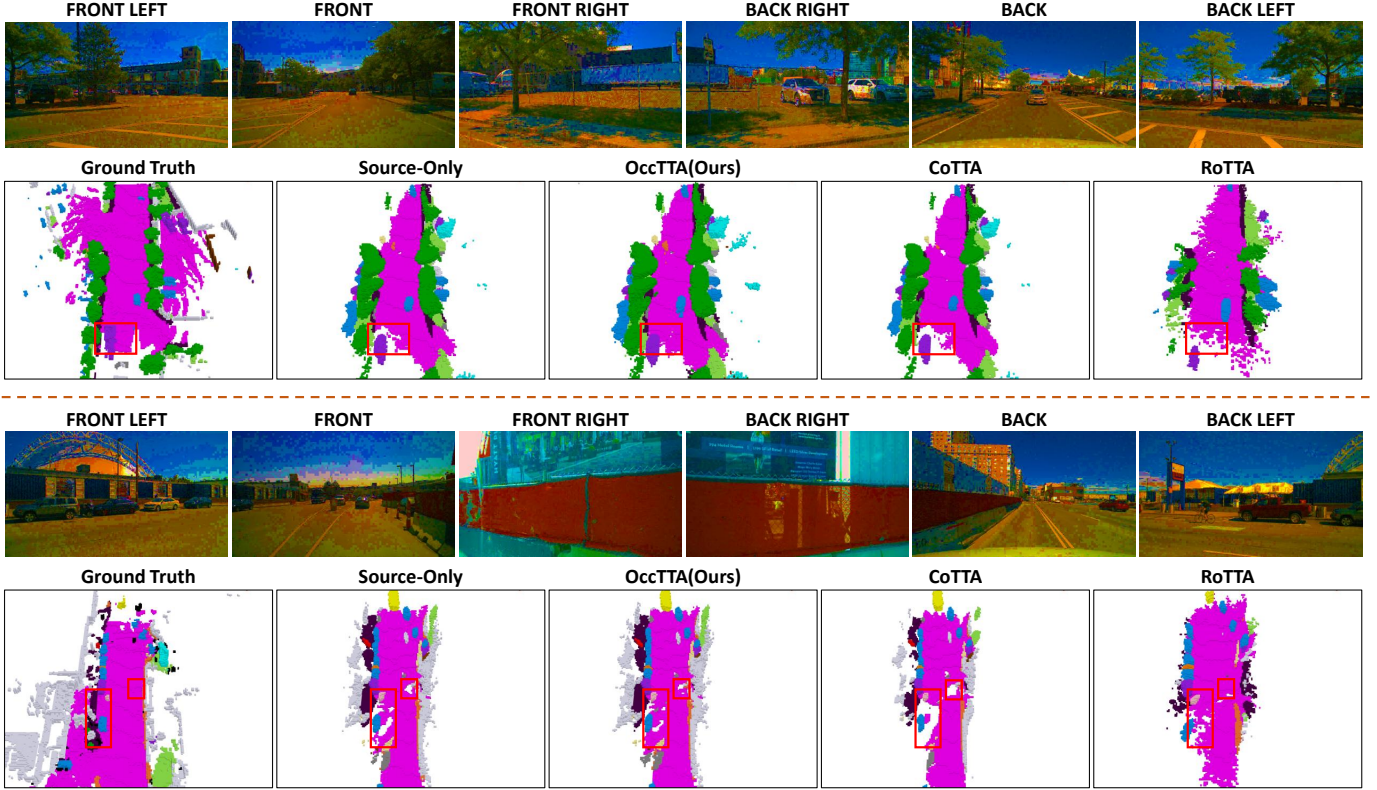


Fig. 8: Qualitative results of baselines and the proposed OccTTA. Failed methods are omitted for clarity and brevity.

noise in individual samples impedes stable model optimization, while aggressive parameter updates in the absence of a mean-teacher framework exacerbate instability; (2) Severe class imbalance further amplifies this bias, aggravating the instability of entropy minimization, which even sharpness-aware or conservative entropy minimization fail to resolve.

B. Qualitative Results

We provide qualitative visualization results from comparison methods and our proposed OccTTA in Fig. 8. It can be observed that the source-only baseline suffers severe performance degradation. While existing TTA methods fail to enhance OCC model performance or even collapse, our approach effectively improves 3D occupancy prediction accuracy under severe domain shifts.

C. Adaptation Performance with Other Backbone.

To further validate the effectiveness of our proposed OccTTA across diverse backbones, we performed additional

experiments with SurroundOcc [10] adopted as an alternative backbone. As presented in Table III, OccTTA consistently attains state-of-the-art performance that outperforms prior TTA methods, demonstrating our method’s consistent effectiveness for different backbones.

D. Ablation Study

Effect of Each Component. As shown in Table IV, compared to the naive solution, incorporating our memory-efficient TTA framework significantly reduces memory overhead while maintaining adaptation efficacy. The adaptive class-balanced loss ensures robust adaptation under severe class imbalance. The forward-backward adaptation strategy significantly accelerates adaptation speed while simultaneously enhancing performance, highlighting its critical role in improving TTA efficiency.

Effectiveness of Forward-backward Adaptation Strategy. As shown in Table V, the integration of efficient forward updates consistently enhances model performance, improving

TABLE IV: Ablation study of each component.

Method	mIoU	Mem.(MB)	Speed.(s/iter)
Naive LayerNorm TTA	9.03	25929	1.70
+Memory-Efficient TTA Framework	9.05	6159	1.40
+Adaptive Class-balanced Loss	9.98	7393	1.68
+Forward-Backward Adaptation Strategy	10.09	7591	1.04

mIoU from 9.98 to 10.09 and 9.97 to 10.09. Selective backward optimization significantly accelerates adaptation speed, decreasing inference time from 1.68s/iter to 1.01s/iter while maintaining comparable performance (mIoU 9.98 vs. 9.97). Notably, the synergistic integration of lightweight forward updates and selective backward optimization achieves both efficiency gains and accuracy improvements simultaneously, attaining a fast inference time of 1.04s/iter alongside a superior mIoU of 10.09.

TABLE V: Ablation study of forward-backward adaptation strategy.

Forward Update	Selective Backward	mIoU	Mem.(MB)	Speed.(s/iter)
-	-	9.98	7393	1.68
✓	-	10.09	7549	1.70
-	✓	9.97	7435	1.01
✓	✓	10.09	7591	1.04

Generalizability of OccTTA. By addressing the unique challenges of OCC tasks while remaining agnostic to specific model architectures, OccTTA can be readily extended to other tasks. To demonstrate this, we conducted additional experiments on the KITTI-C dataset for 3D object detection task using MonoFlex [33]. As shown in Table VI, OccTTA achieves state-of-the-art performance superior to recent work MonoTTA [33] (**32.77 vs. 34.69mAP**) while simultaneously reducing memory overhead by **27.2%** and accelerating adaptation speed by **11.8%**. This demonstrates our method’s consistent effectiveness beyond OCC tasks while maintaining efficiency advantages.

TABLE VI: Generalization to monocular 3D object detection.

Method	mAP	Mem.(MB)	Speed.(ms/iter)
Source-only	20.55	4244	26
MonoTTA [33]	32.77	15150	68
OccTTA(Ours)	34.69	11022	60

Architecture Ablation of Lite-boost Units. We conducted additional ablation experiments to analyze different design choices for the lite-boost units. As shown in Table VII below, varying the number of Linear-ReLU-LN blocks has little impact on adaptation performance. Since a single Linear-ReLU-LN block may have limited adaptive capacity, leading to slightly weaker performance in the table, while three Linear-ReLU-LN blocks introduce extra memory overhead, we opted for two Linear-ReLU-LN blocks to balance performance and memory consumption.

Adaptation Performance with Single Corruption Type. We conduct additional experiments to further evaluate adaptation performance with a single corruption type. As shown in Table VIII below, even when focusing on adapting to a single corruption type, existing TTA methods still struggle to improve

TABLE VII: Ablation study of design choices for the lite-boost units.

Architecture of lite-boost units	Average mIoU	Mem.(MB)
One Linear-ReLU-LN block	10.08	7450
Two Linear-ReLU-LN blocks	10.09	7591
Three Linear-ReLU-LN blocks	10.09	8006

model performance, which underscores the unique challenges of test-time adaptation for OCC models. In contrast, our method achieves robust adaptation under class imbalance conditions, consistently delivering state-of-the-art performance. This demonstrates its distinctive effectiveness in adapting OCC models to domain-shifted data. Furthermore, the speed and memory efficiency of our method further highlight its superiority.

TABLE VIII: Adaptation performance with single corruption type.

Method	Gaussian	Defocus	Frost	Contrast
Source-only	2.77	5.80	8.21	5.80
TENT [23]	0.01	0.02	0.05	0.03
MEMO [26]	0.01	0.01	0.01	0.01
SAR [28]	1.36	3.04	4.43	3.04
CoTTA [25]	2.72	5.74	7.56	5.64
RoTTA [27]	2.82	5.93	8.22	6.05
OccTTA(Ours)	3.51	7.29	9.68	7.63

Ablation on Warm-up Initialization for Lite-Boost Units.

We further investigate the impact of the warm-up stage for initializing the lite-boost units by comparing it with a zero-initialized setting. In this setting, the newly introduced lite-boost units are initialized with zero weights, so that the residual branch starts from a neutral state and is updated online during test-time adaptation. As shown in Table IX, zero initialization achieves an average mIoU of 6.78, slightly lower than the 6.92 obtained with warm-up initialization, indicating that warm-up provides a useful but modest initialization benefit. Compared with the source-only baseline, OccTTA with zero initialization still yields consistent gains, increasing the average mIoU from 5.65 to 6.78. These results demonstrate that OccTTA remains effective across different lite-boost initialization settings, including cases where source-data warm-up is unavailable.

TABLE IX: Ablation on warm-up initialization for lite-boost units.

Initialization	Gaussian	Defocus	Frost	Contrast	Avg. mIoU
Source-only	2.77	5.80	8.21	5.80	5.65
Zero initialization	3.58	7.01	9.25	7.29	6.78
Warm-up initialization	3.51	7.14	9.50	7.51	6.92

Ablation on the Moving-Average Design. To isolate the contribution of the moving-average design, we compare the full loss with a variant that directly uses the current class-wise confidence for reweighting. As shown in Table X, removing the moving average slightly decreases the mIoU from 10.09 to 10.05, indicating that it provides a modest benefit by stabilizing the difficulty estimation. Meanwhile,

the variant without moving average still maintains competitive performance, suggesting that the adaptive class-balanced loss remains effective under domain shifts.

TABLE X: Ablation on the moving-average design in the adaptive class-balanced loss.

Variant	mIoU
Loss w/o moving average	10.05
Loss w/ moving average	10.09

Ablation on k-NN Memory Bank. To further analyze the cost and necessity of the k-NN memory bank, we conduct two ablation studies on memory-bank scalability and interval-wise adaptation behavior. First, we evaluate the scalability of the k-NN memory bank by varying the number of source and target slots per class. As shown in Table XI, a small memory bank with 650 slots per class improves mIoU from 9.97 to 10.03, with only 78 MB extra memory and 0.02 s/iter additional runtime. Increasing the capacity to our default setting of 1300 slots per class further improves mIoU to 10.09, while the overhead remains moderate. A larger memory bank with 1950 slots per class only brings a marginal gain to 10.10 but incurs higher cost. Therefore, we adopt 1300 slots per class as a practical trade-off between accuracy, memory consumption, and speed. Second, we analyze the necessity of the memory bank. As shown in Table XII, the k-NN classifier (Θ_{knn}) rapidly responds to domain shifts through lightweight forward updates, achieving superior performance in early adaptation stages compared to the original classifier (Θ_c). By fusing predictions from both classifiers ($\Theta_{combine}$), our method accelerates responsiveness to domain shifts and attains enhanced overall performance.

TABLE XI: Scalability of the k-NN memory bank.

Source slots/class	Target slots/class	Total slots/class	Pred. Mem. (MB)	Pred. Speed (s/iter)	Pred. mIoU
0	0	0	7435	1.01	9.97
500	150	650	7513	1.03	10.03
1000	300	1300	7591	1.04	10.09
1500	450	1950	7669	1.05	10.10

TABLE XII: Comparisons of mIoU at corresponding iteration intervals between Θ_{knn} , Θ_c and $\Theta_{combine}$. Experiments are conducted on the nuScenes-C validation set with Gaussian noise corruption.

Classifier	0–100	100–500	500–1000	1000–2000	2000–6000
Θ_{knn}	4.09	4.85	3.95	3.02	3.35
Θ_c	3.95	4.77	3.90	3.00	3.43
$\Theta_{combine}$	4.00	4.81	3.92	3.01	3.44

Comparison with LoRA-based Adaptation. To clarify the distinction between lite-boost units and LoRA, we compare OccTTA with a LoRA-based variant, where LoRA with rank 4 is applied to the same OCC modules under identical test-time adaptation settings. As shown in Table XIII, the LoRA-based variant achieves 8.93 mIoU with 8364 MB memory and 1.18 s/iter, while OccTTA obtains 10.09 mIoU with 7591 MB memory and 1.04 s/iter. These results show that our lite-boost design not only improves adaptation performance but also

maintains better memory efficiency and adaptation speed. We attribute this advantage to two factors. First, LoRA performs low-rank updates on individual linear layers and therefore cannot directly provide a module-level residual correction across the whole cross-view attention or refinement module. Applying LoRA to these modules requires retaining activations for multiple internal linear projections, leading to higher memory and computational overhead in TTA. Second, test-time adaptation is performed without explicit labels and is thus vulnerable to noisy self-supervised gradients. LoRA perturbs internal backbone projections, which can lead to unstable representation drift. In contrast, our lite-boost units preserve the frozen pretrained pathway and learn only a lightweight normalized residual branch, making the adaptation more stable under unsupervised OCC-TTA scenarios.

TABLE XIII: Comparison with LoRA-based test-time adaptation. LoRA with rank 4 is applied to the same modules under identical TTA settings.

Method	mIoU	Mem.(MB)	Speed(s/iter)
LoRA ($r = 4$)	8.93	8364	1.18
OccTTA	10.09	7591	1.04

Sensitivity to the Reset Threshold ϵ' . We further analyze the influence of the reset threshold ϵ' in continual TTA. As shown in Table XIV, OccTTA maintains stable performance around the default threshold. In particular, varying ϵ' from 1.5 to 2.5 only causes marginal mIoU fluctuations, from 10.07 to 10.09. These results demonstrate that OccTTA is not sensitive to ϵ' within a reasonable range around the default setting.

TABLE XIV: Sensitivity analysis of the reset threshold ϵ' .

Reset threshold ϵ'	1.0	1.5	2.0	2.5	3.0
Avg. mIoU	10.01	10.07	10.09	10.08	10.02

Failure Diagnosis of Self-Entropy Minimization Baselines.

To better understand the collapse of self-entropy-minimization baselines, we conduct a detailed failure diagnosis using TENT as a representative case, recording per-frame gradient statistics and class distributions under Gaussian noise corruption. As shown in Table XV, TENT rapidly converges to a trivial empty-class solution. The inherent sparsity of OCC labels induces a strong empty-class bias. During entropy-minimization-based adaptation, this bias is further amplified. The predicted empty ratio increases to 99.92% at frame 17 and 99.996% at frame 26. By frame 50, the prediction degenerates to a single active class, i.e., the empty class. Meanwhile, the gradient norm drops from 0.1384 at frame 0 to 0.0026 at frame 50 and 0.0005 at frame 100, indicating that the adaptation signal quickly vanishes once the model enters the over-confident all-empty solution. These observations explain why entropy-based baselines such as TENT, MEMO, and COME are prone to collapse in our OCC setting: entropy minimization reinforces the dominant empty-class prediction under severe occupancy class imbalance, rather than providing a corrective adaptation signal.

We further considered fairer variants by applying adaptive class balancing to TENT. As shown in Table XVI, this

TABLE XV: Per-frame failure diagnosis of TENT under Gaussian noise.

Frame	Pred. Empty (%)	GT Empty (%)	Pred. Active Cls.	Grad. Norm	Ent. Loss
0	98.365	94.869	10	0.1384	0.1065
10	99.446	94.704	9	0.0773	0.0616
17	99.922	94.211	8	0.0548	0.0435
26	99.996	93.951	4	0.0252	0.0305
50	100.000	92.310	1	0.0026	0.0223
100	100.000	94.949	1	0.0005	0.0220

variant slightly improves over vanilla TENT but still remains far below effective adaptation performance, with only 0.11 average mIoU. This result shows that adaptive class balancing mitigates the empty-class bias, but is not sufficient by itself to stabilize standalone entropy minimization under severe OCC class imbalance.

TABLE XVI: TENT variant with adaptive class balancing.

Method	Gaussian	Defocus	Frost	Contrast	Avg. mIoU
TENT [23]	0.01	0.00	0.00	0.00	0.00
TENT w/ adaptive class-balancing	0.15	0.09	0.10	0.08	0.11

Relation to long-tailed learning. Our adaptive class-balanced loss is related to class re-balancing in long-tailed learning, but differs from standard long-tailed techniques in the TTA context. Conventional methods such as class re-weighting [50], margin adjustment [51], logit adjustment [52]–[54], and decoupled classifier retraining [55] typically rely on labeled training data with known class frequencies, and use this static frequency prior to correct head/tail decision-boundary bias during offline supervised training. In contrast, test-time adaptation is source-free and label-free, where the target class distribution is unknown, temporally changing, and only observable through noisy predictions. Directly using pseudo-label frequencies can be unreliable because OCC predictions are dominated by the empty class and rare-class pseudo-labels may contain substantial noise. Therefore, our loss does not depend on static class frequencies. Instead, it dynamically estimates class-wise adaptation difficulty from online confidence statistics and reweights the adaptation objective accordingly, stabilizing TTA under severe empty-class dominance and non-stationary target streams.

Theoretical Analysis of Selective Backward Optimization. Let ϕ_t denote the learnable parameters of the lite-boost units and $\mathcal{L}_t(\phi_t)$ denote the adaptation objective at time step t . The selective backward optimization can be formulated as an event-triggered gradient descent:

$$\phi_{t+1} = \phi_t - \eta s_t \nabla_{\phi} \mathcal{L}_t(\phi_t), \quad s_t = \mathbf{1}\{\cos(y_t, o_{t-1}) < \epsilon\}. \quad (16)$$

For local analysis, assume that within a locally stationary segment, $\mathcal{L}_t$ can be approximated by a local objective $\mathcal{L}$ that is L -smooth and lower bounded. When the learning rate satisfies $0 < \eta < 2/L$, each triggered backward update satisfies the standard descent condition:

$$\mathcal{L}(\phi_{t+1}) \leq \mathcal{L}(\phi_t) - \eta \left(1 - \frac{L\eta}{2}\right) s_t \|\nabla \mathcal{L}(\phi_t)\|_2^2. \quad (17)$$

Therefore, with $K_T = \sum_{t=1}^T s_t$, the average squared gradient norm over triggered updates is bounded by $\mathcal{O}(1/K_T)$. This indicates that selective backward optimization preserves the local descent behavior of gradient-based adaptation while reducing redundant backward computation. Moreover, for skipped

frames, $s_t = 0$ implies $\cos(y_t, o_{t-1}) \geq \epsilon$. For normalized output statistics $\hat{y}_t$ and $\hat{o}_{t-1}$, we have

$$\|\hat{y}_t - \hat{o}_{t-1}\|_2 = \sqrt{2 - 2\cos(\hat{y}_t, \hat{o}_{t-1})} \leq \sqrt{2(1 - \epsilon)}. \quad (18)$$

Thus, the threshold ϵ explicitly controls the maximum discrepancy between the current and reference statistics allowed for skipping backward computation. Under the local Lipschitz assumption that the adaptation gradient varies smoothly with the output statistics, the skipped gradient discrepancy is bounded by $\mathcal{O}(\sqrt{1 - \epsilon})$ plus the residual local distribution drift. Therefore, a large threshold ϵ makes the selective update conservative: backward computation is skipped only when the current frame is close to the reference statistics, providing a threshold-controlled stability bound for online adaptation.

VI. LIMITATIONS

Although OccTTA improves the robustness of OCC models under domain shifts in a fast and memory-efficient manner, it still has limitations. The adaptation performance under severe corruptions remains constrained by the quality of pseudo-labels. While the proposed adaptive class-balanced loss alleviates this issue to some extent, designing specific adaptation objectives, such as geometry-aware or temporal-consistency-based losses, remains a promising direction for further improving test-time adaptation in OCC. In addition, real-device profiling on automotive SoCs or embedded platforms remains an important direction for future deployment-oriented evaluation.

VII. CONCLUSION

In this paper, we presented OccTTA, a novel test-time adaptation method specifically designed to address the unique challenges of 3D semantic occupancy prediction in autonomous driving. We identified two key issues: memory efficiency for deployment on resource-constrained devices and the difficulty of rapidly adapting to dynamic scenes to ensure fast prediction. Our proposed method overcomes these challenges by leveraging a memory-efficient TTA framework based on lite-boost units and a forward-backward adaptation strategy to ensure fast and memory-efficient adaptation during inference. Furthermore, we introduced an adaptive class-balanced loss to address the prevalent class imbalance issue in OCC, ensuring robust adaptation to challenging classes. Extensive experiments on the nuScenes-C benchmark demonstrate the effectiveness and efficiency of our OccTTA in real-world testing scenarios.

VIII. ACKNOWLEDGMENTS

This work is supported in part by the National Key R&D Program of China (NO. 2024YFB3908503 and NO. 2024YFB3908500), in part by the National Natural Science Foundation of China (No. 62322608), and in part by the PCL Major Key Project under Grant PCL2025A17-2. The authors would like to thank the National Supercomputer Center in Guangzhou for providing high performance computational resources.

REFERENCES

- [1] J. Huang, G. Huang, Z. Zhu, Y. Ye, and D. Du, "Bevdet: High-performance multi-camera 3d object detection in bird-eye-view," *arXiv preprint arXiv:2112.11790*, 2021.
- [2] Y. Li, Z. Ge, G. Yu, J. Yang, Z. Wang, Y. Shi, J. Sun, and Z. Li, "Bevdepth: Acquisition of reliable depth for multi-view 3d object detection," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 37, no. 2, 2023, pp. 1477–1485.
- [3] Y. Wang, V. C. Guizilini, T. Zhang, Y. Wang, H. Zhao, and J. Solomon, "Detr3d: 3d object detection from multi-view images via 3d-to-2d queries," in *Conference on Robot Learning*. PMLR, 2022, pp. 180–191.
- [4] S. Wang, Y. Liu, T. Wang, Y. Li, and X. Zhang, "Exploring object-centric temporal modeling for efficient multi-view 3d object detection," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2023, pp. 3621–3631.
- [5] X. Zhu, X. Cao, Z. Dong, C. Zhou, Q. Liu, W. Li, and Y. Wang, "Nemo: Neural map growing system for spatiotemporal fusion in bird's-eye-view and bdd-map benchmark," *arXiv preprint arXiv:2306.04540*, 2023.
- [6] B. Zhou and P. Krähenbühl, "Cross-view transformers for real-time map-view semantic segmentation," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 13 760–13 769.
- [7] L. Peng, Z. Chen, Z. Fu, P. Liang, and E. Cheng, "Bevsegformer: Bird's eye view semantic segmentation from arbitrary camera rigs," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2023, pp. 5935–5943.
- [8] J. Phillion and S. Fidler, "Lift, splat, shoot: Encoding images from arbitrary camera rigs by implicitly unprojecting to 3d," in *Computer Vision—ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XIV 16*. Springer, 2020, pp. 194–210.
- [9] C. Yang, Y. Chen, H. Tian, C. Tao, X. Zhu, Z. Zhang, G. Huang, H. Li, Y. Qiao, L. Lu *et al.*, "Bevformer v2: Adapting modern image backbones to bird's-eye-view recognition via perspective supervision," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 17 830–17 839.
- [10] Y. Wei, L. Zhao, W. Zheng, Z. Zhu, J. Zhou, and J. Lu, "Surroundocc: Multi-camera 3d occupancy prediction for autonomous driving," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 21 729–21 740.
- [11] Y. Huang, W. Zheng, Y. Zhang, J. Zhou, and J. Lu, "Tri-perspective view for vision-based 3d semantic occupancy prediction," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2023, pp. 9223–9232.
- [12] Y. Zhang, Z. Zhu, and D. Du, "Occformer: Dual-path transformer for vision-based 3d semantic occupancy prediction," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 9433–9443.
- [13] Q. Ma, X. Tan, Y. Qu, L. Ma, Z. Zhang, and Y. Xie, "Cotr: Compact occupancy transformer for vision-based 3d occupancy prediction," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 19 936–19 945.
- [14] J. Hou, X. Li, W. Guan, G. Zhang, D. Feng, Y. Du, X. Xue, and J. Pu, "Fastocc: Accelerating 3d occupancy prediction by fusing the 2d bird's-eye view and perspective view," in *2024 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2024, pp. 16 425–16 431.
- [15] H. Liu, Y. Chen, H. Wang, Z. Yang, T. Li, J. Zeng, L. Chen, H. Li, and L. Wang, "Fully sparse 3d occupancy prediction," in *European Conference on Computer Vision*. Springer, 2024, pp. 54–71.
- [16] Y. He, W. Chen, T. Xun, and Y. Tan, "Real-time 3d occupancy prediction via geometric-semantic disentanglement," *arXiv preprint arXiv:2407.13155*, 2024.
- [17] Y. Wang, Y. Chen, X. Liao, L. Fan, and Z. Zhang, "Panoocc: Unified occupancy representation for camera-based 3d panoptic segmentation," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2024, pp. 17 158–17 168.
- [18] M. Pan, J. Liu, R. Zhang, P. Huang, X. Li, H. Xie, B. Wang, L. Liu, and S. Zhang, "Renderocc: Vision-centric 3d occupancy prediction with 2d rendering supervision," in *2024 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2024, pp. 12 404–12 411.
- [19] W. Gan, N. Mo, H. Xu, and N. Yokoya, "A simple framework for 3d occupancy estimation in autonomous driving," *arXiv preprint arXiv:2303.10076*, 2023.
- [20] Y. Huang, W. Zheng, B. Zhang, J. Zhou, and J. Lu, "Selfocc: Self-supervised vision-based 3d occupancy prediction," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 19 946–19 956.
- [21] C. Zhang, J. Yan, Y. Wei, J. Li, L. Liu, Y. Tang, Y. Duan, and J. Lu, "Ocnerf: Advancing 3d occupancy prediction in lidar-free environments," *IEEE Transactions on Image Processing*, 2025.
- [22] Y. Huang, W. Zheng, Y. Zhang, J. Zhou, and J. Lu, "Gaussianformer: Scene as gaussians for vision-based 3d semantic occupancy prediction," in *European Conference on Computer Vision*. Springer, 2024, pp. 376–393.
- [23] D. Wang, E. Shelhamer, S. Liu, B. Olshausen, and T. Darrell, "Tent: Fully test-time adaptation by entropy minimization," in *International Conference on Learning Representations*.
- [24] S. Goyal, M. Sun, A. Raghunathan, and J. Z. Kolter, "Test time adaptation via conjugate pseudo-labels," *Advances in Neural Information Processing Systems*, vol. 35, pp. 6204–6218, 2022.
- [25] Q. Wang, O. Fink, L. Van Gool, and D. Dai, "Continual test-time domain adaptation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 7201–7211.
- [26] M. Zhang, S. Levine, and C. Finn, "Memo: Test time robustness via adaptation and augmentation," *Advances in neural information processing systems*, vol. 35, pp. 38 629–38 642, 2022.
- [27] 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.
- [28] S. Niu, J. Wu, Y. Zhang, Z. Wen, Y. Chen, P. Zhao, and M. Tan, "Towards stable test-time adaptation in dynamic wild world," in *The Eleventh International Conference on Learning Representations*.
- [29] S. Choi, S. Yang, S. Choi, and S. Yun, "Improving test-time adaptation via shift-agnostic weight regularization and nearest source prototypes," in *European Conference on Computer Vision*. Springer, 2022, pp. 440–458.
- [30] S. Jung, J. Lee, N. Kim, A. Shaban, B. Boots, and J. Choo, "Cafa: Class-aware feature alignment for test-time adaptation," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 19 060–19 071.
- [31] H. Lim, B. Kim, J. Choo, and S. Choi, "Ttn: A domain-shift aware batch normalization in test-time adaptation," in *11th International Conference on Learning Representations, ICLR 2023*, 2023.
- [32] S. Niu, J. Wu, Y. Zhang, Y. Chen, S. Zheng, P. Zhao, and M. Tan, "Efficient test-time model adaptation without forgetting," in *International conference on machine learning*. PMLR, 2022, pp. 16 888–16 905.
- [33] H. Lin, Y. Zhang, S. Niu, S. Cui, and Z. Li, "Monotta: Fully test-time adaptation for monocular 3d object detection," in *European Conference on Computer Vision*. Springer, 2024, pp. 96–114.
- [34] J. Wu, X. Luo, Y. Zhou, L. Wu, G. Wang, and S. Zhang, "A3-tta: Adaptive anchor alignment test-time adaptation for image segmentation," *IEEE Transactions on Image Processing*, vol. 34, pp. 8511–8522, 2025.
- [35] Y. Liu, C. Huang, Y. Xu, X. Cao, and J. Wang, "Towards efficient test time adaptation with hierarchical distribution alignment," *IEEE Transactions on Image Processing*, 2025.
- [36] A. Zhang, T. Yu, L. Bai, J. Tang, Y. Guo, Y. Ruan, Y. Zhou, and Z. Lu, "Cola: Context-aware language-driven test-time adaptation," *IEEE Transactions on Image Processing*, 2025.
- [37] T. Lim, J.-W. Hwang, and K. Lee, "When and where to reset matters for long-term test-time adaptation," in *The Fourteenth International Conference on Learning Representations*.
- [38] Z. Liao, P. Wei, S. Chen, H. Wang, and Z. Ren, "Stocc: Sparse spatial-temporal cascade renovation for 3d occupancy and scene flow prediction," in *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2025, pp. 1516–1526.
- [39] J. Park, C. Peng, W. Zheng, K. Kitani, W. Zhan *et al.*, "S2go: Streaming sparse gaussian occupancy," in *International Conference on Learning Representations*, vol. 2026, 2026, pp. 62 911–62 932.
- [40] Q. Zhang, Y. Bian, X. Kong, P. Zhao, and C. Zhang, "Come: Test-time adaption by conservatively minimizing entropy," in *International Conference on Learning Representations*, 2025.
- [41] Z. Hu, Y. Hu, X. Li, S. TANG, and L. DUAN, "Beyond entropy: Region confidence proxy for wild test-time adaptation," in *Forty-second International Conference on Machine Learning*, 2025. [Online]. Available: <https://openreview.net/forum?id=QwxjjaGBUo>
- [42] S. Maharana, B. Zhang, L. Karlinsky, R. Feris, and Y. Guo, "Batchclip: Bimodal online test-time adaptation for clip," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2025, pp. 1569–1579.
- [43] L. Zhou, M. Ye, S. Li, N. Li, X. Zhu, L. Deng, H. Liu, and Z. Lei, "Bayesian test-time adaptation for vision-language models," in *Proceedings of the Computer Vision and Pattern Recognition Conference*, 2025, pp. 29 999–30 009.

- [44] A. Li, L. Zhuang, X. Long, H. Li, and S. Wang, “Flatness guided test-time adaptation for vision-language models,” in *The Fourteenth International Conference on Learning Representations*.
- [45] S. Dhar, J. Guo, J. Liu, S. Tripathi, U. Kurup, and M. Shah, “A survey of on-device machine learning: An algorithms and learning theory perspective,” *ACM Transactions on Internet of Things*, vol. 2, no. 3, pp. 1–49, 2021.
- [46] Y. Nie, C. Fang, L. Cheng, L. Lin, and G. Li, “Adapting object size variance and class imbalance for semi-supervised object detection,” in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 37, no. 2, 2023, pp. 1966–1974.
- [47] C. Fang, L. Cheng, Y. Mao, D. Zhang, Y. Fang, G. Li, H. Qi, and L. Jiao, “Separating noisy samples from tail classes for long-tailed image classification with label noise,” *IEEE Transactions on Neural Networks and Learning Systems*, 2023.
- [48] H. Caesar, V. Bankiti, A. H. Lang, S. Vora, V. E. Liong, Q. Xu, A. Krishnan, Y. Pan, G. Baldan, and O. Beijbom, “nusenes: A multimodal dataset for autonomous driving,” in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 11 621–11 631.
- [49] D. Hendrycks and T. Dietterich, “Benchmarking neural network robustness to common corruptions and perturbations,” in *International Conference on Learning Representations*.
- [50] Y. Cui, M. Jia, T.-Y. Lin, Y. Song, and S. Belongie, “Class-balanced loss based on effective number of samples,” in *2019 IEEE/CVF conference on computer vision and pattern recognition (CVPR)*. IEEE, 2019, pp. 9260–9269.
- [51] K. Cao, C. Wei, A. Gaidon, N. Arechiga, and T. Ma, “Learning imbalanced datasets with label-distribution-aware margin loss,” *Advances in neural information processing systems*, vol. 32, 2019.
- [52] A. K. Menon, S. Jayasumana, A. S. Rawat, H. Jain, A. Veit, and S. Kumar, “Long-tail learning via logit adjustment,” *arXiv preprint arXiv:2007.07314*, 2020.
- [53] Y. Tao, J. Sun, H. Yang, L. Chen, X. Wang, W. Yang, D. Du, and M. Zheng, “Local and global logit adjustments for long-tailed learning,” in *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*. IEEE, 2023, pp. 11 749–11 758.
- [54] E. Zhang, C. Geng, C. Li, and S. Chen, “Dynamic learnable logit adjustment for long-tailed visual recognition,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 34, no. 9, pp. 7986–7997, 2024.
- [55] B. Kang, S. Xie, M. Rohrbach, Z. Yan, A. Gordo, J. Feng, and Y. Kalantidis, “Decoupling representation and classifier for long-tailed recognition,” *arXiv preprint arXiv:1910.09217*, 2019.



Zijian He received a B.E. degree from the School of Mathematical Sciences, South China Normal University, Guangzhou, China, in 2017 and an M.S. degree from the School of Information, Xiamen University, Xiamen, China, in 2020. He is currently pursuing a Ph.D. degree in computer science at Sun Yat-sen University. His research interests include multimedia and computer vision.



Qiqing Lao received the B.E. degree from the School of Computer Science and Engineering, Sun Yat-sen University, Guangzhou, China, in 2024, where she is currently pursuing an M.S. degree in computer science. Her research interests include machine learning and computer vision.



Yipeng Qin received the Ph.D. degree from National Centre for Computer Animation (NCCA), Bournemouth University, UK, in 2017. He is currently a Senior Lecturer (Associate Professor) with the School of Computer Science and Informatics, Cardiff University, UK. His research interests include machine learning, computer vision, computer graphics, and human-computer interaction. Dr. Qin has published extensively in top-tier venues, including CVPR, ICCV, ECCV, NeurIPS, ICML, and SIGGRAPH. His work has received several recognitions, including the SIGGRAPH 2025 Best Paper Award and CVPR 2024 Best Paper Award Candidate. He has served as an Area Chair for major conferences such as ICML, NeurIPS and BMVC, and as a reviewer or program committee member for leading journals and conferences including TPAMI, IJCV, SIGGRAPH, CVPR, ICCV, and ECCV.



Peixin Chen received the B.E. degree from the School of Computer Science and Engineering, Sun Yat-sen University, Guangzhou, China, in 2024, where he is currently pursuing an M.S. degree in computer science. His research interests include computer vision and machine learning.



Jie Feng (Senior Member, IEEE) received the B.S. degree from Chang'an University, Xi'an, China, in 2008, and the Ph.D. degree from Xidian University, Xi'an, in 2014. She is currently a Professor with the Laboratory of Intelligent Perception and Image Understanding, Xidian University. Her research interests include remote sensing image processing, deep learning, and machine learning.



Wei Zhang received a Ph.D. degree from the University of Hong Kong in 2017. He is currently a senior Engineer in the Department of Computer Vision Technology at Baidu Inc. His current research interests include 2D/3D object detection, segmentation, and tracking. He is dedicated to the development of computer vision models for autonomous driving and intelligent transportation systems.



Guanbin Li (Member, IEEE) received the Ph.D. degree from The University of Hong Kong in 2016. He is currently a full professor with the School of Computer Science and Engineering, Sun Yat-sen University. He has authored and coauthored more than 150 papers in top-tier academic journals and conferences. His current research interests include computer vision, image processing, and deep learning. He was a recipient of the ICCV 2019 Best Paper Nomination Award. He serves as an Area Chair for the conference of VISAPP. He has been serving as a reviewer for numerous academic journals and conferences, such as IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, International Journal of Computer Vision, IEEE TRANSACTIONS ON IMAGE PROCESSING, IEEE TRANSACTIONS ON MULTIMEDIA, IEEE TRANSACTIONS ON CYBERNETICS, CVPR, ICCV, ECCV, and NeurIPS.