Cross-Architecture Adaptation: Cloud-Edge Continual Test-Time Adaptation with Dynamic Sampling and Heterogeneous Distillation

Supplementary Material

A. Detailed Implementation Settings

Due to space constraints in the main text, we provide further implementation details of our Cross-Architecture Adaptation (CAA) framework here. Both the cloud teacher (ViT-B) and the edge student (ResNet-18) are optimized using Stochastic Gradient Descent (SGD) with a momentum of 0.9.

For the edge-side distillation updates, we employ a dynamic learning rate schedule to prevent over-fitting to noisy pseudolabels. We initialize the learning rate at 0.03. To maintain stability, we monitor the cross-entropy loss $\mathcal{L}_{CE}$ on the distillation batches. When $\mathcal{L}_{CE}$ falls below the threshold $\tau_{ce} = 0.23$ for T consecutive mini-batches, indicating that the local adaptation has converged for the current distribution, we decay the learning rate to 0.00025.

B. Additional Experimental Results

B.1. Full Results on ImageNet-C (Severity Level 4)

In the main paper, we reported the comprehensive results for the most severe distribution shifts (Severity Level 5). In Table S1, we provide the complete breakdown of the adaptation accuracy across all 15 corruption types for Severity Level 4. Our CAA framework consistently outperforms existing state-of-the-art methods across the majority of corruptions.

B.2. Extensive Generalization Settings

To validate the generalization ability of CAA across diverse network architectures and datasets, we conducted additional experiments:

- **Model Diversity:** When deploying a heavier ResNet-50 on the edge while keeping the Cloud ViT-B unchanged, CAA achieves 48.4% accuracy on ImageNet-C, outperforming the SOTA ETA (46.8%).
- **Homogeneous Architectures:** Our method is not strictly limited to heterogeneous setups. When utilizing a Cloud ViT-B and an Edge TinyViT (homogeneous Transformer structures), CAA achieves 46.6%, which is superior to the SOTA CEMA (43.5%). This demonstrates that CAA effectively coordinates inductive biases even in isomorphic scenarios.
- **Dataset Generalization:** We further evaluated CAA on the ImageNet-R dataset, which features diverse renditions of ImageNet classes. CAA achieves 28.3%, outperforming CEMA (27.6%).

C. Hyperparameter Sensitivity

The hyperparameters introduced in our framework demonstrate strong robustness. We specifically analyze the crucial balancing parameter α (which weights sample uncertainty against representativeness in the MDCS module) and the cross-entropy threshold τ_{ce} used for the learning rate schedule.

α	0.1	0.3	**0.5**	0.7	0.9
Acc(%)	39.4	40.1	**41.2**	40.8	40.4
τ_{ce}	0.10	0.20	**0.23**	0.30	0.40
Acc(%)	40.5	41.0	**41.2**	40.8	40.6

Table S2. Sensitivity analysis of hyperparameters α and τ_{ce} on ImageNet-C.

As shown in Table S2, we control for a comparable number of selected samples to ensure a fair comparison. CAA consistently outperforms state-of-the-art methods across different α values, maintaining high stability with minimal

	Noise			Blur				Weather				Digital				
Severity Level=4	Gauss.	Shot	Impul.	Defoc.	Glass	Motion	Zoom	Snow	Frost	Fog	Brit.	Contr.	Elastic	Pixel	JPEG	Avg.
ResNet18 (baseline)	7.8	6.2	6.5	18.6	12.1	16.2	22.1	17.5	21.9	28.4	57.9	14.2	39.3	26.4	43.6	22.6
• BN Adaptation†	29.8	24.9	28.2	26.8	25.7	33.0	39.2	31.4	35.0	50.5	62.2	44.6	54.4	49.0	49.0	38.9
• CoTTA	29.2	25.6	27.1	21.2	23.3	34.7	41.4	32.8	35.4	53.2	62.5	43.8	54.9	51.6	50.1	39.1
• ETA	41.4	41.9	42.0	32.9	37.0	40.3	44.9	39.7	39.9	**53.7**	59.5	50.4	56.3	54.1	53.1	45.6
• CEMA	42.2	43.4	43.0	32.0	39.0	43.0	47.5	40.9	42.1	**53.7**	60.0	49.2	**57.4**	55.1	**54.9**	46.3
• CoLA	41.2	42.0	42.2	**34.4**	38.4	40.8	45.0	38.7	40.2	53.0	58.9	**50.6**	56.6	54.4	53.5	45.3
• CAA(Ours)	**43.3**	**44.4**	**44.5**	33.0	**40.3**	**43.5**	**48.0**	**42.9**	**43.1**	**53.7**	**60.4**	49.0	57.3	**55.7**	54.7	**47.6**

Table S1. Comparisons with state-of-the-art methods on ImageNet-C (severity level 4) regarding **Accuracy (%)**. We adopt ViT-B as the foundation model and ResNet18 as the edge model. † denotes the TTA method that does not require any backward propagation and can be locally executed in edge devices.

deviation within the range of $[0.3, 0.7]$. Furthermore, the performance remains stable across various τ_{ce} values, peaking at $\tau_{ce} = 0.23$.

D. Computational Efficiency

Although MDCS requires a dual forward pass to estimate uncertainty, the resulting precise sample selection drastically reduces the computational overhead associated with backpropagation. As a result, the per-epoch adaptation time on ImageNet-C for CAA is 49.6 ms, which is highly comparable to a full-sample update strategy without any selection (46.3 ms). Furthermore, our online clustering strategy involves a feature- and centroid-assignment process with a computational complexity of $\mathcal{O}(N \times d)$ (where $d = 512$), ensuring that the edge-side computation remains minimal.