Supplementary Material

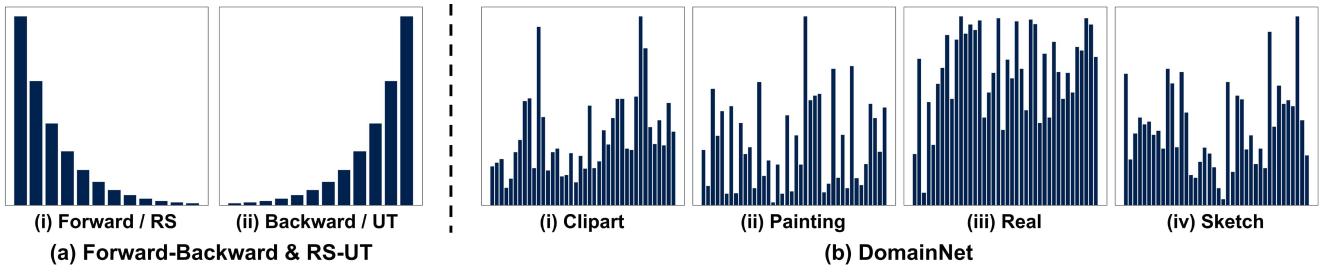

Figure 6. **Visualization of label distributions in datasets.** (a) shows the illustrations of forward or reversely-unbalanced source (RS) setting and backward or unbalanced target (UT) setting. In specific, forward and backward are used in CIFAR-10-C, CIFAR-100-C, and ImageNet-C. In addition, RS and UT are utilized in VisDA-C and OfficeHome. (b) shows the natural label shift of DomainNet.

A. Implementation Details

In this section, we introduce further information regarding the datasets, along with the implementation details for the baseline test-time-adaptation (TTA) methods and the label shift adapter.

A.1. Datasets

Fig. 6 illustrates the label distributions for the datasets utilized in our experiments. As depicted in Fig. 6 (a), 'forward' and 'RS' represent long-tailed label distributions, with class order corresponding to the training label distribution. Conversely, 'backward' and 'UT' indicate a reversed class order.

In the forward and backward settings, the imbalance ratios for CIFAR-10-C, CIFAR-100-C, and ImageNet-C are configured to 10, 25, and 100. We adjust the label distribution by reducing the number of images per class based on the specified imbalance ratio. For VisDA-C, The imbalance ratio is set to 100 for both training and test datasets. Furthermore, we utilize an imbalanced version of OfficeHome created by the previous research [43].

Fig. 6 (b) shows the label distributions of DomainNet, in which existing label shifts are significant enough. The superior performance of our method on DomainNet demonstrates its ability to handle label shifts that arise in real-world scenarios.

A.2. Details of Baselines

We carry out the experiments using the official implementations of the baseline models. We provide additional details regarding the implementation specifics, including hyperparameters. Note that the batch size for test-time adaptation is configured to 64 for fair comparisons. For simplicity, we present the hyperparameters in the following sequence: **{CIFAR-10-C, CIFAR-100-C, ImageNet-C, VisDA-C, OfficeHome, DomainNet}** for test-time adaptation baselines. In instances where hyperparameters are not separately described for each dataset, the same values are employed across all datasets.

Source. Different from the previous TTA studies, we employ long-tailed datasets in our research. To mitigate model bias towards the majority classes, we utilize a balanced softmax [39], which is a prominent method for long-tailed recognition. Formally, the balanced softmax is expressed as:

$$\mathcal{L}_{\text{bal}} = - \sum_{(x_i, y_i) \sim \mathcal{D}_s} y_i \log \sigma(\hat{y}_i + \log(\pi_s)),$$

where π_s represents the frequency of the training classes, and σ denotes the softmax function.

Table 9 describes the hyperparameters utilized for training on source domain datasets. We select the hyperparameters for VisDA-C, OfficeHome, and DomainNet in accordance with the imbalanced source-free domain adaptation study [22]. As described in the main manuscript, we utilize pre-trained ResNet-50 and ResNet-101 on ImageNet, when conducting the experiments on VisDA-C, OfficeHome, and DomainNet. Moreover, the learning rate for the feature extractor and the classifier is set to 0.1×LR and LR, respectively, when training the model on VisDA-C, OfficeHome, and DomainNet. All experiments are conducted using NVIDIA RTX A5000 GPU.

BN Stats. BN stats [41] utilizes test batch statistics instead of running statistics within batch normalization layers.

PseudoLabel. In accordance with previous studies [20, 46],

Src Data	Tgt Data	Model	Optim.	Scheduler	Epoch	Batch	WD	Momentum	LR
CIFAR-10-LT	CIFAR-10-C	ResNet-18	SGD	CosineAnneal	200	128	5e-4	0.9	0.1
CIFAR-100-LT	CIFAR-100-C	ResNet-18	SGD	CosineAnneal	200	128	5e-4	0.9	0.1
ImageNet-LT	ImageNet-C	ResNeXt-50	SGD	Manual	90	64	2e-4	0.9	0.1
VisDA-C (RS)	VisDA-C (UT)	ResNet-101	SGD	-	15	40	1e-3	0.9	1e-3
OfficeHome (RS)	OfficeHome (UT)	ResNet-50	SGD	-	50	40	1e-3	0.9	1e-2
DomainNet	DomainNet	ResNet-50	SGD	-	20	40	1e-3	0.9	1e-2

Table 9. **Hyperparameters for training the model with source domain data.** Src Data and Tgt Data denote source domain and target domain datasets, respectively. Optim. indicates the optimizer. WD and LR denote the weight decay and learning rate for training. The manual scheduler for ImageNet-LT is to decay the learning rate at 60 and 80 epochs.

Model Architecture				Forward-LT			Uni.	Backward-LT			Avg.
γ_h	β_h	ΔW	Δb	50	25	10	1	10	25	50	
✓				51.20	49.30	46.06	37.36	27.28	23.75	21.84	36.69
	✓			48.79	42.45	32.10	15.21	22.52	24.09	25.14	30.04
		✓		51.32	49.36	46.11	37.17	27.06	23.56	21.71	36.61
			✓	**52.50**	**50.19**	**46.64**	**37.51**	28.95	25.56	24.06	37.92
✓	✓			52.09	49.48	45.43	35.92	**29.60**	**26.82**	**26.25**	37.94
		✓	✓	51.93	49.78	46.43	37.18	27.71	24.28	22.57	37.12
✓	✓	✓	✓	52.06	49.71	46.03	36.84	29.29	26.33	25.50	**37.97**

Table 10. **Ablation study on architecture design of label shift adapter using CIFAR-100-LT and CIFAR-100-C.**

we update the affine parameters in the batch normalization layers using the hard pseudo labels. The learning rate is set to {1e-3, 1e-3, 2.5e-4, 5e-5, 5e-5, 1e-3} for each respective dataset, following the hyperparameters of TENT [46].

ONDA. Online domain adaptation (ONDA) [32] modifies the batch normalization statistics for target domains using a batch of target data through an exponential moving average. We set the update frequency $N = 10$ and the decay of the moving average $m = 0.1$, adhering to the default values of the original paper.

TENT. Test entropy minimization (TENT) [46] optimizes the affine parameters of batch normalization layers via entropy minimization. The learning rate is configured to {1e-3, 1e-3, 2.5e-4, 5e-5, 5e-5, 1e-3} for each dataset. We referred to the official implementation for hyperparameter selection.

LAME. Laplacian adjusted maximum-likelihood estimation (LAME) [4] alters the output probability of the classifier. Following the authors' implementation, we set the kNN affinity matrix with the value of k as 5.

CoTTA. Continual test-time adaptation (CoTTA) [47] adapts the model to accommodate continually evolving target domains by employing a weight-averaged teacher model, data augmentations, and stochastic restoring. CoTTA incorporates three hyperparameters: augmentation confidence threshold p_{th}, restoration factor p, and the decay of EMA m. p and m are set to 0.01 and 0.999, respectively. Additionally, p_{th} is configured to { 0.92, 0.72, 0.01, 0.01, 0.01, 0.01 }. Given that the authors do not provide the hyperparameters for VisDA-C, OfficeHome, and DomainNet, we fine-tune the appropriate hyperparameters for them.

Algorithm 1 Training Process of Label Shift Adapter

Require: Dataset $\mathcal{D}_s = \{(x_i, y_i)\}_{i=1}^n$. A pre-trained model f. A label shift adapter $\mathcal{G}_\phi$.

Initialize the parameters ϕ randomly
for $k = 1$ to K **do**
  $\mathcal{B} \leftarrow$ SampleMiniBatch$(\mathcal{D}, m)$ ▷ a mini-batch of m examples
  $\pi, \tau \leftarrow$ Sample$(\{\pi_s, u, \bar{\pi}_s\}, \{\tau_{\pi_s}, \tau_u, \tau_{\bar{\pi}_s}\})$ ▷ sample τ matching π
  $\mathcal{L}(\mathcal{G}_\phi) \leftarrow \frac{1}{m} \sum_{(x,y)\in\mathcal{B}} \mathcal{L}_{gla}((x, y, \pi); f, \mathcal{G}_\phi)$
  $\mathcal{G}_\phi \leftarrow \mathcal{G}_\phi - \eta \nabla_\theta \mathcal{L}(\mathcal{G}_\phi)$ ▷ one SGD step
end for

NOTE. Non-i.i.d. test-time adaptation (NOTE) [11] comprises two components: (i) Instance-aware batch normalization (IABN), and (ii) Prediction-balanced reservoir sampling (PBRS). In accordance with the original paper, we substitute the batch normalization layers with IABN layers before pre-training the source models. Two hyperparameters are associated with IABN: soft-shrinkage width α and EMA momentum m. The values of α are configured as { 4, 4, 8, 8, 8, 8 }, while m is set to { 0.01, 0.01, 0.1, 0.1, 0.1, 0.1 }. The memory size of PBRS is set to 64, equal to the batch size. In our experiments, we incorporate our label shift adapter into the models using IABN layers.

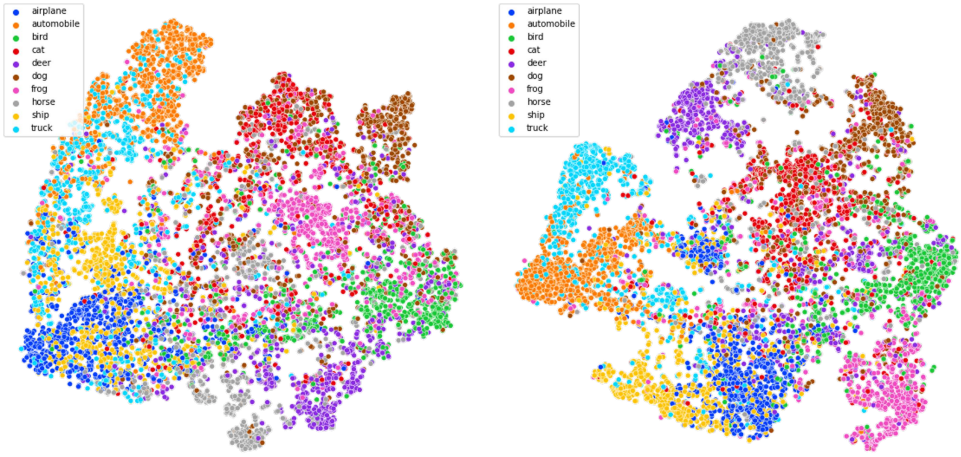

Figure 7. **T-SNE visualizations of (a) TENT and (b) IABN+Ours.** We visualize the feature map h obtained from the Gaussian noise corruption in the CIFAR-10-C uniform test dataset. The number of training samples is large in the order of the classes in the legend.

Src Method	TTA Method	Forward-LT			Uni.	Backward-LT			Avg.
		50	25	10	1	10	25	50	
Cross Entropy	Source	40.20	37.26	33.18	22.20	12.19	8.95	7.24	23.03
	BN Stats	48.12	45.44	40.17	26.17	13.67	9.69	7.80	27.30
	ONDA	48.49	45.80	41.16	27.66	15.17	11.02	8.98	28.33
	PseudoLabel	48.76	45.26	39.09	19.84	11.32	8.19	6.44	25.56
	TENT	49.17	45.21	38.42	15.32	9.99	7.44	5.67	24.46
	LAME	38.17	35.17	31.22	20.44	11.02	7.93	6.30	21.46
	CoTTA	32.83	29.35	25.72	14.75	7.38	4.94	3.67	16.95
	NOTE	46.41	44.10	40.49	29.30	15.77	11.87	9.84	28.26
	IABN	46.43	43.92	39.80	25.35	14.71	11.13	9.27	27.23
	+Ours	**53.20**	**50.77**	**46.26**	**32.34**	**18.56**	**14.07**	**11.74**	**32.42**
Balanced Sampling	Source	34.88	32.54	29.12	20.29	11.80	9.09	7.51	20.75
	BN Stats	45.07	42.93	39.10	28.67	17.82	13.97	11.88	28.49
	ONDA	44.79	42.60	39.01	29.20	18.62	14.65	12.70	28.80
	PseudoLabel	47.08	44.66	40.45	26.98	17.31	13.59	11.23	28.76
	TENT	48.28	45.64	41.07	24.04	16.88	13.50	11.30	28.67
	LAME	32.88	30.44	27.08	18.48	10.48	7.89	6.39	19.09
	CoTTA	30.38	28.68	25.58	17.32	10.74	7.96	6.49	18.16
	NOTE	46.62	44.27	40.64	30.42	16.99	12.77	10.32	28.86
	IABN	46.85	44.29	40.40	27.20	16.12	12.05	9.88	28.12
	+Ours	**51.32**	**49.18**	**44.99**	**31.92**	**19.64**	**15.11**	**13.00**	**32.16**
Classifier Re-Training	Source	40.17	37.47	33.59	23.23	13.42	10.18	8.51	23.80
	BN Stats	48.33	45.60	40.82	27.49	15.13	11.16	9.11	28.23
	ONDA	48.33	45.77	41.57	28.96	16.78	12.50	10.63	29.22
	PseudoLabel	49.10	45.89	40.19	21.01	12.74	9.50	7.66	26.58
	TENT	49.70	46.19	39.79	16.59	11.28	8.93	6.92	25.63
	LAME	38.22	35.41	31.60	21.48	12.28	9.13	7.52	22.23
	CoTTA	31.93	29.23	25.91	15.26	8.18	5.55	4.33	17.20
	NOTE	45.96	44.04	41.14	31.68	18.52	14.66	12.67	29.81
	IABN	46.11	44.04	40.53	27.61	17.28	13.71	11.94	28.75
	+Ours	**53.11**	**50.86**	**46.66**	**34.04**	**20.77**	**16.41**	**14.12**	**33.71**

Table 11. **Ablation study on the source pre-trained model using CIFAR-100-LT and CIFAR-100-C.**

A.3. Details of Label Shift Adapter

Model Architecture. We utilize the same model architecture for the label shift adapter across all datasets. The proposed label shift adapter consists of two fully-connected (FC) layers and a ReLU activation function, structured as FC-ReLU-FC. Furthermore, the label shift adapter is partitioned into two neural networks producing (γ_h, β_h) and (ΔW, Δb). As described in the main manuscript, the label shift adapter takes $m^\intercal\pi \in \mathbb{R}^1$ and produces (γ_h, β_h) and (ΔW, Δb) in each respective neural network. The hidden layer size in the label shift adapter is configured to 100.

Details of Label Shift Adapter. We provide the algorithm of the training process for the label shift adapter as a pseudo-code in Algorithm 1. The primary objective of the label shift adapter is to learn the relationship between π and adaptive parameters by selecting appropriate τ based on sampled π within generalized logit adjusted loss [33, 1] function. Increasing τ results in decision boundary shifting away from the minority class towards the majority class. Consequently, instead of sampling batches differently based on π, we sample π and τ iteratively, as described in Algorithm 1. This enables the label shift adapter to optimize its parameters in accordance with input label distributions (*e.g.*, π and $\hat{\mathcal{Y}}_t$), thereby producing suitable parameter adjustments.

During the training of the label shift adapter, we sample the label distribution π from three types of label distributions: $\{\pi_s, u, \bar{\pi}_s\}$. For each sampled label distribution, we select the appropriate $\tau \subset \{\tau_{\pi_s}, \tau_u, \tau_{\bar{\pi}_s}\}$, with the hyperparameter τ corresponding to each π. Different τ values are employed for each dataset. We set τ to $\{1, -1.5, 3\}$, $\{1, 0, -2\}$, $\{1, 0, -2\}$, $\{1, 0, -2\}$, $\{1, -1, -3\}$, and $\{1, 0, -2\}$, for CIFAR-10-LT, CIFAR-100-LT, ImageNet-LT, VisDA-C, OfficeHome, and DomainNet, respectively.

The mapping vector m maps the label distribution's vector to the scalar of the imbalance degree. We set the range of m from -1 to 1, with the values increasing proportionally to the data count rank of each class. This technique enables the adapter to effectively utilize the degree of imbalance as an input, circumventing the challenges associated with complex label spaces encountered when using π directly.

While training the label shift adapter, we employ the same optimizer and batch size as those employed for training the source models. The learning rate is set to 1e-3 for all datasets. Moreover, we train the label shift adapter for $\{200, 200, 30, 15, 50, 20\}$ epochs.

During inference, the momentum hyperparameter α for target label distribution estimation is configured to 0.1. For learnable parameters in the test-time adaptation process, we only update affine parameters in normalization layers by following TENT [46] and IABN [11]. Unlike TENT, we freeze the top layers and update the affine parameters of the layer in the remaining shallow layers, inspired by previous work [7, 36]. Specifically, for ResNet, including four layer groups (layer 1, 2, 3, 4), we only freeze layer4 in CIFAR-10-C, CIFAR-100-C, and ImageNet-C. In other datasets, there is no significant difference in performance, so all affine parameters are trained. When estimating the label distribution on ImageNet-C, we utilize only the top-3 probability to update the estimated label distribution $\hat{\mathcal{Y}}_t$. Empirically, we discovered that it is effective to consider only top-k when the number of classes is particularly large.

B. Further Analysis on Label Shift Adapter

Ablation Study on Architecture Design. We examine the model architecture design for the proposed label shift adapter. The label shift adapter produces four types of outputs: γ_h, β_h, ΔW, and Δb. Table 10 presents the ablation study for each component. Interestingly, even when only Δb is employed, the performance is quite good. However, we observed that as the degree of the label shift increases, the performance of using only Δb declines. Moreover, utilizing γ_h and β_h only also yields impressive results, indicating that appropriately shiting the feature map h is effective in addressing the label shifts. We choose the architecture design of the label shift adapter that achieves the best average accuracy, indicating that the final model generally performs well across a variety of label distributions.

T-SNE Visualization. To further substantiate the effectiveness of our method, we visualize the feature map h using t-SNE by extracting h during test-time adaptation. As illustrated in Fig. 7, our method shows a more well-separated representation space in a class-wise manner compared to TENT. Notably, it is evident that the minority classes (*e.g.*, horse and truck) are not well divided in the representation space of TENT. In contrast, our method integrating into IABN layers enhances class-discriminability.

Ablation Study on Source Model. In the main manuscript, we employ the balanced softmax to reduce the model bias towards the majority classes. To further validate the effectiveness of the proposed method, we apply our method to several source pre-trained models utilizing different training strategies. We employ three types of techniques: (i) Cross-entropy loss, (ii) Balanced sampling, (iii) Classifier re-training [18], where the feature extractor is trained using cross-entropy loss, and then the classifier is randomly re-initialized and re-trained using class-balanced sampling. Table 11 demonstrates that our method effectively handles the label shifts, regardless of the source pre-trained models. Moreover, these results indicate that existing long-tailed recognition methods can be combined with our method to further reduce the model bias towards the majority classes in source domain data.

Ablation Study on π. As described in the main manuscript, we sampled three kinds of label distributions for π during

Num.	F50	F25	F10	U	B10	B25	B50	Avg.
3	**52.06**	**49.71**	**46.03**	36.84	**29.29**	26.33	25.50	**37.97**
5	50.91	48.82	45.41	36.90	29.28	**26.37**	**25.51**	37.60
7	51.13	48.99	45.52	36.96	29.24	26.25	25.45	37.64
∞	51.62	49.36	45.85	**37.09**	29.12	26.06	25.03	37.73

Table 12. Ablation study on the number of π for training label shift adapter using CIFAR-100-C. Num. denotes the number of π for training the adapter. F, U, and B indicate forward, uniform, and backward distributions, respectively. We chose three label distributions.

	DELTA	ISFDA	TENT+Ours
VISDA-C	50.10	61.02	**72.97**

Table 13. Comparison with additional baselines in test-time adaptation setting.

	Method	F50	F25	F10	U	B10	B25	B50	Avg.
CIFAR10	SAR+GN	57.22	57.20	57.07	57.12	61.84	63.06	64.37	59.70
	SAR+BN	78.63	76.28	71.82	53.28	34.99	28.60	25.18	52.68
	Ours+IABN	**80.58**	**78.62**	**75.26**	**63.34**	**68.54**	**70.07**	**71.64**	**72.58**
CIFAR100	SAR+GN	9.09	9.59	10.23	14.05	18.93	20.46	21.70	14.86
	SAR+BN	49.44	47.04	43.39	32.18	20.22	16.24	13.96	31.78
	Ours+IABN	**52.06**	**49.71**	**46.03**	**36.84**	**29.29**	**26.33**	**25.50**	**37.97**

Table 14. Comparison with SAR using CIFAR-10-C and CIFAR-100-C in TTA setting. F, U, and B denote forward, uniform, and backward, respectively. GN and BN indicate group and batch normalization, respectively.

training label shift adapter. Regarding the effect of sampling different numbers of π, Table 12 indicates that such variations have negligible impact on performance. Specifically, in this experiment, we interpolate three distributions (*i.e.*, π_s, u, $\bar{\pi}_s$) and τ to train the label shift adapter when different numbers of π are utilized.

C. Additional Experiments

Comparison with Baselines Related to Label Shifts. We've compared two baselines in Table 7, which have the capability of handling label shifts. We compare additional baselines, DELTA [53] and ISFDA [22], which address covariate and label shifts simultaneously. Although ISFDA requires several epochs for adapting the source models, we conduct the experiments in the test-time adaptation setting for a fair comparison. Table 13 demonstrates that our method is superior to baselines significantly in the VISDA-C dataset. ISFDA, a domain adaptation model, exhibits limitations in its suitability for online learning during inference. Since DELTA only focuses on class imbalances in the target domain, it lacks the ability to handle imbalances in the source domain. In contrast, our method successfully addresses the imbalance in both source and target domains in the test-time adaptation setting.

Comparison with Recent TTA Baseline. We compare recent test-time adaptation baseline, sharpness-aware and reliable entropy minimization (SAR) [36]. SAR proposes an optimizer and analyzes normalization layers to resolve imbalances in the target domain. However, it is important to note that our work addresses imbalances in both the source and target domains. Table 14 demonstrates that our method outperforms both SAR+GN and SAR+BN significantly. Moreover, it is a viable option to integrate our method with SAR method.

D. Domain-wise Results

Table 15, 16, 17 show the average classification accuracy on CIFAR-10-C, CIFAR-100-C, and ImageNet-C, shown per domain. To compute the accuracy of each domain, we calculate the average performances of Forward50, Forward25, Forward10, Uniform, Backward10, Backward25, and Backward50, as described in the main manuscript. These results demonstrate that our method consistently enhances performance across various domains.

Method	*Gaussian*	*Shot*	*Impulse*	*Defocus*	*Glass*	*Motion*	*Zoom*	*Snow*	*Frost*	*Fog*	*Brightness*	*Contrast*	*Elastic*	*Pixelate*	*JPEG*	**Avg**
Source	23.06	26.94	19.59	46.89	41.06	45.63	49.42	58.30	45.31	45.72	69.85	21.23	57.60	60.14	65.45	45.08
BN Stats	49.18	50.52	46.29	58.11	45.05	56.14	55.96	52.32	50.65	53.60	59.11	54.93	51.63	54.15	52.12	52.65
ONDA	50.70	51.66	47.68	59.80	46.49	57.45	57.78	53.81	52.36	55.14	61.52	54.76	53.60	56.49	54.19	54.23
PseudoLabel	46.87	48.76	44.47	55.96	43.87	53.89	53.46	49.96	48.70	50.91	56.34	52.57	49.29	51.88	50.16	50.47
LAME	17.97	22.75	15.43	44.74	40.36	42.40	47.08	61.30	48.62	45.20	67.84	20.33	55.40	61.36	64.59	43.69
CoTTA	51.69	53.11	50.31	55.80	47.28	54.31	54.88	52.60	52.18	52.81	58.30	49.66	52.12	55.32	54.39	52.98
NOTE	54.48	56.22	53.24	68.20	48.64	64.87	65.09	65.56	64.43	64.08	73.33	67.59	60.24	66.95	67.26	62.68
TENT	46.41	48.29	43.38	53.82	42.42	52.22	51.57	48.92	47.51	49.81	55.06	50.62	47.90	50.58	49.20	49.18
+ Ours	51.66	53.55	48.66	60.74	46.94	58.77	57.37	55.04	53.48	56.00	62.05	57.68	54.11	57.48	55.07	55.24
IABN	54.77	56.48	53.25	68.24	48.39	64.53	64.89	65.63	64.44	64.60	73.79	67.24	60.32	67.81	67.17	62.77
+ Ours	**68.72**	**69.54**	**64.94**	**77.48**	**61.47**	**76.24**	**75.52**	**72.06**	**72.83**	**74.53**	**79.69**	**79.08**	**69.66**	**74.25**	**72.69**	**72.58**

Table 15. **Domain-wise results on CIFAR-10-C.**

Method	*Gaussian*	*Shot*	*Impulse*	*Defocus*	*Glass*	*Motion*	*Zoom*	*Snow*	*Frost*	*Fog*	*Brightness*	*Contrast*	*Elastic*	*Pixelate*	*JPEG*	**Avg**
Source	14.42	16.45	9.08	22.55	23.14	25.65	25.82	27.37	22.98	18.57	34.51	5.84	33.03	17.47	36.19	22.21
BN Stats	27.59	28.51	26.20	36.89	28.61	34.53	36.44	29.40	28.81	28.98	36.66	29.22	33.37	33.78	32.18	31.41
ONDA	27.53	28.77	26.17	36.85	29.07	34.37	36.91	29.71	29.09	29.38	36.97	27.77	33.89	34.00	33.00	31.57
PseudoLabel	27.44	28.54	25.59	34.92	27.78	32.83	34.56	28.58	27.40	28.58	34.96	26.01	31.58	32.86	31.38	30.20
LAME	13.41	15.73	7.66	20.93	22.30	24.79	24.63	27.21	22.39	17.07	33.62	4.48	32.41	15.62	35.66	21.19
CoTTA	30.01	30.85	28.45	34.77	30.64	34.04	35.64	30.92	30.10	28.65	36.09	24.30	33.54	35.58	34.00	31.84
NOTE	24.17	25.64	18.62	35.73	28.08	36.89	37.48	34.91	33.95	29.13	41.47	33.93	36.29	32.01	36.13	32.30
TENT	27.50	28.49	25.28	33.98	26.89	32.12	33.36	28.00	26.79	27.78	34.05	25.20	31.01	32.25	30.53	29.55
+ Ours	30.95	32.21	**28.77**	38.60	30.58	36.06	38.24	32.30	30.74	31.52	38.34	30.36	34.75	36.19	34.86	33.63
IABN	24.54	25.79	18.92	35.50	28.00	36.80	37.58	34.97	34.20	29.02	41.39	33.99	36.17	32.09	36.29	32.35
+ Ours	**33.65**	**34.37**	28.17	**41.86**	**33.62**	**41.08**	**41.79**	**37.82**	**38.36**	**34.77**	**43.51**	**42.54**	**39.18**	**39.98**	**38.76**	**37.97**

Table 16. **Domain-wise results on CIFAR-100-C.**

Method	*Gaussian*	*Shot*	*Impulse*	*Defocus*	*Glass*	*Motion*	*Zoom*	*Snow*	*Frost*	*Fog*	*Brightness*	*Contrast*	*Elastic*	*Pixelate*	*JPEG*	**Avg**
Source	5.72	5.88	4.82	15.48	11.17	17.98	18.73	21.25	15.21	28.04	44.67	28.63	39.16	29.56	34.77	21.40
BN Stats	29.27	28.90	25.46	25.59	22.94	33.14	32.91	28.37	25.15	40.06	45.27	40.19	43.59	41.84	39.74	33.49
ONDA	29.15	28.88	25.33	25.49	22.74	32.73	32.96	28.18	25.04	40.12	45.46	39.69	43.60	42.02	39.72	33.41
PseudoLabel	31.25	30.93	29.00	27.72	25.80	34.36	33.64	30.20	25.39	40.33	44.09	39.56	42.78	41.43	39.65	34.41
LAME	5.56	5.73	4.67	15.33	11.03	17.92	18.67	21.19	15.16	28.01	44.64	28.61	39.12	29.49	34.75	21.33
CoTTA	30.93	30.36	27.47	27.28	24.95	34.42	33.56	30.01	26.44	40.62	44.79	40.58	43.32	41.83	40.05	34.44
NOTE	31.34	30.83	29.26	27.39	24.66	35.67	32.70	33.34	28.33	39.52	46.55	44.97	43.67	41.38	41.59	35.41
TENT	29.29	28.94	25.84	25.68	22.94	32.11	32.37	27.17	23.50	39.43	43.94	38.18	42.30	40.66	39.02	32.76
+ Ours	32.38	32.39	28.77	29.07	26.21	35.70	**35.81**	31.44	27.86	**43.14**	48.56	43.00	46.59	44.92	43.59	36.63
IABN	31.47	30.86	29.28	27.37	24.66	35.69	32.77	33.35	28.37	39.54	46.65	45.03	43.72	41.40	41.60	35.45
+ Ours	**34.28**	**34.00**	**32.18**	**30.25**	**27.14**	**38.92**	35.53	**36.48**	**31.58**	42.50	**49.95**	**48.35**	**46.93**	**44.95**	**44.87**	**38.53**

Table 17. **Domain-wise results on ImageNet-C.**