DO NOT CROSS THE BOUNDARY: WITHIN-CLASS WHITENING FOR TEST-TIME ADAPTATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Test-time adaptation (TTA) recalibrates a deployed model under domain shift using only unlabeled test data, and recent methods have substantially improved the robustness of such models. However, most existing TTA methods adjust only class-agnostic statistics and overlook the distortion of the feature distribution within each class. Concretely, domain shift elongates the covariance of each class, but existing TTA methods leave it elongated even after adaptation, causing some samples to cross the decision boundary and thus remain misclassified. To mitigate this, we propose Within-Class Whitening (WCW), a plug-in method that whitens the penultimate features with the class-averaged within-class covariance, without backpropagation or pre-computed source statistics. We analyze the family of affine feature transformations and show that WCW makes the within-class covariance isotropic, which lowers an upper bound on the probability of crossing the decision boundary under domain shift. We also identify that restricting the whitening scope to the within-class component avoids the implicit injection of a uniform class prior caused by the rearrangement of the class means. Extensive experiments on various datasets, backbones, and adaptation scenarios show that WCW consistently improves representative parameter-updating TTA methods, outperforms existing TTA plug-ins in most cases, and substantially stabilizes the adaptation procedure.

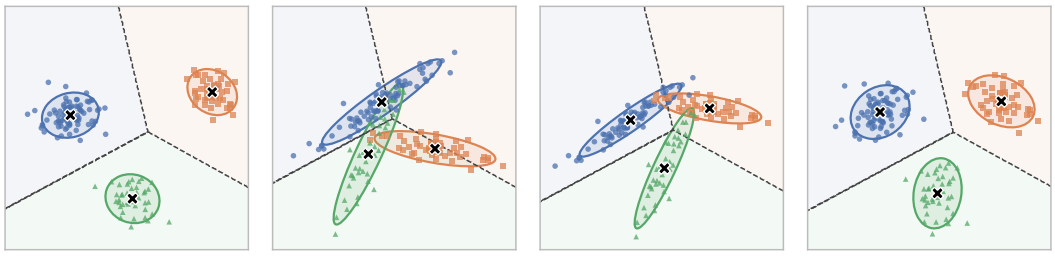
1 INTRODUCTION

Deep neural networks (DNNs) have recently made significant progress, but deployed DNN models often suffer performance degradation due to the discrepancy between the source and target domains, known as domain shift (Hendrycks & Dietterich, 2019; Hendrycks et al., 2021). Therefore, this discrepancy has become a central challenge in machine learning, motivating extensive research on adaptation frameworks such as unsupervised domain adaptation (Sun et al., 2016; Li et al., 2018; Roy et al., 2019) and test-time training (Sun et al., 2020; Gandelsman et al., 2022). More recently, test-time adaptation (TTA) has emerged as a compelling alternative, since it adapts a model with unlabeled test data alone (Wang et al., 2021; Lee et al., 2024).

Several TTA methods address domain shift by adjusting the feature statistics (Mirza et al., 2023; Ma et al., 2026; You et al., 2025; Murphy et al., 2026). These approaches rest on the assumption that restoring the statistics of the feature distribution to those of the source would substantially compensate for the performance degradation (Schneider et al., 2020). However, most of these methods deal only with class-agnostic statistics, such as the global mean and the total covariance, and pay little attention to the distortion of the distribution within each class. In practice, domain shift not only disturbs the whole distribution of the features but also deforms their scatter within each class.

Figures 1(a) and 1(b) illustrate the distortion. On the source domain (Figure 1(a)), the features of each class scatter about their own mean, and the class means are well separated from one another. The scatter of each class remains moderate, and its features lie far from the decision boundaries. Under domain shift (Figure 1(b)), the class means are displaced from their source positions, and the scatter within each class is distorted into an elongated shape. As the scatter is stretched along certain

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.



(a) Source domain (b) Target domain (c) Existing TTA (d) WCW (Ours)

Figure 1: **Two-dimensional schematic of feature variation.** Each color and marker shape denotes a class, ellipses outline the scatter of each class, crosses mark the class means, and dashed lines represent the decision boundaries.

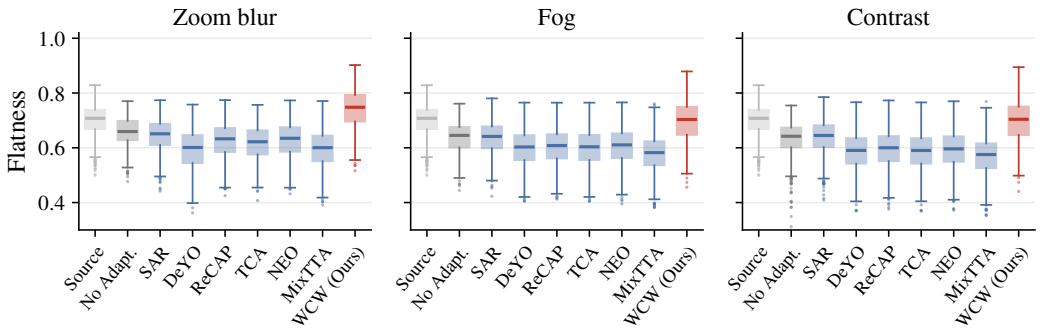


Figure 2: **Spectral flatness of the covariance of each class under adaptation (ViT-B/16, ImageNet-C).** Each box shows the distribution over the 1,000 classes, with the median and the quartiles marked and the whiskers extending to 1.5 times the interquartile range. The results for the other corruptions are reported in Appendix H.

directions, a portion of it can cross a decision boundary, causing the corresponding features to be misclassified.

To examine how existing TTA methods treat this distortion, we measure the spectral flatness (Gray & Markel, 1974) of the covariance of each class during adaptation. The flatness is defined as the ratio of the geometric mean of the eigenvalues to their arithmetic mean and equals one when the covariance is isotropic. Under domain shift, the flatness of most classes falls below its value on the source domain, as Figure 2 shows. Across six adaptation methods—SAR (Niu et al., 2023), DeYO (Lee et al., 2024), ReCAP (Hu et al., 2025), TCA (You et al., 2025), NEO (Murphy et al., 2026), and MixTTA (Jung et al., 2026), where the last three are plug-ins applied on top of DeYO—the flatness of most classes stays below the source level even after adaptation, and the covariance of each class remains far from isotropic. These measurements suggest that existing TTA methods leave the distortion of the within-class covariance largely unaddressed.

Motivated by this observation, we propose Within-Class Whitening (WCW), a feature-transformation-based approach that modifies the within-class covariance directly. It whitens the penultimate features based on the within-class covariance, so that the distorted scatter becomes closer to isotropic and fewer samples spill into the decision regions of adjacent classes. Figures 1(c) and 1(d) contrast existing adaptation with the proposed method. After adaptation with existing TTA methods (Figure 1(c)), the class means have moved back close to their source positions, but the scatter within each class remains distorted, so a portion of the features still crosses a decision boundary. With WCW (Figure 1(d)), the scatter of each class becomes nearly isotropic and the features lie on their own side of the boundaries. As Figure 2 shows, WCW raises the flatness above its value under any of the existing TTA methods.

Our analysis further shows that restricting the whitening to the within-class covariance avoids interfering with the arrangement of the class means. Since this arrangement is coupled with the predicted

class distribution, excluding the class-mean covariance prevents the whitening from distorting the class prior. Notably, WCW requires neither backpropagation nor any offline pass over source data and can be integrated with existing parameter-updating TTA methods in a fully online manner. Our experiments show that this simple whitening strategy suffices to obtain consistent accuracy gains across the evaluated settings and stabilizes adaptation in scenarios where it is often fragile.

2 Related work

Test-time adaptation via model update. TTA improves the robustness of a deployed model to domain shift using only unlabeled test data. Early TTA methods mostly update the affine parameters of normalization layers to control the first- and second-order moments of intermediate features. For instance, Tent (Wang et al., 2021) updates normalization parameters by minimizing the prediction entropy, and later methods keep this update rule while augmenting it with complementary components: EATA (Niu et al., 2022) filters individual samples by their entropy and proposes an anti-forgetting regularizer, SAR (Niu et al., 2023) rejects high-gradient samples and flattens the loss surface, DeYO (Lee et al., 2024) introduces another filtering criterion to detect spurious correlations, ReCAP (Hu et al., 2025) employs a region confidence proxy, and COME (Zhang et al., 2025a) treats the prediction as a Dirichlet opinion with an explicit uncertainty mass to mitigate overconfidence. In parallel, another line of work replaces or complements entropy minimization with self-supervised learning. This line includes pseudo-label self-training (Wang et al., 2023), contrastive learning (Chen et al., 2022), temporal ensembling (Wang et al., 2022; Döbler et al., 2023), and transformation invariance (Zhang et al., 2022; Nguyen et al., 2023). Regardless of the adaptation objective, these methods depend on updating the model parameters. In contrast, WCW transforms the features directly and requires no model update, so it operates without backpropagation, even on a frozen model.

Domain adaptation via feature statistics. Domain shift can also be addressed by adjusting the feature statistics themselves. AdaBN (Li et al., 2018) replaces the statistics of batch normalization with those of the target domain. CORAL (Sun et al., 2016) matches second-order statistics across domains. DWT (Roy et al., 2019) replaces batch normalization with whitening layers that decorrelate the intermediate features of each domain with its own batch covariance. Several TTA methods bring this strategy to test time by aligning the statistics of the target features with those of the source. TTAC (Su et al., 2022) models the target features as class-wise Gaussians and aligns them with source anchors. CAFA (Jung et al., 2023) aligns target features with class-conditional source Gaussians through a Mahalanobis objective. ActMAD (Mirza et al., 2023) aligns the mean and variance of intermediate activations across multiple layers to those of the training set. FOA (Niu et al., 2024) learns an input prompt with a derivative-free evolution strategy whose fitness function employs the prediction entropy and the discrepancy between target and source activation statistics. PEA (Ma et al., 2026) characterizes domain shift as a translation, scaling, and rotation of the embeddings and progressively realigns every block with the source statistics through a whitening-and-coloring transformation. However, most of the methods above rely on information from the source domain, whether its data or its pre-computed statistics, which is often unavailable in practical deployments.

Source-free adjustment of feature statistics. To overcome the dependence on information from the source domain, several approaches exploit the test stream alone. T3A (Iwasawa & Matsuo, 2021) replaces the linear classifier with class prototypes of pseudo-labeled test features without backpropagation. TCA (You et al., 2025) uses high-certainty test samples as a surrogate source and fits a linear map that aligns the feature covariance with that of these samples, and NEO (Murphy et al., 2026) recenters the features with a running mean. ADAPT (Zhang et al., 2025b) and DOTA (Han et al., 2025) target vision-language models and complement their zero-shot classifiers with a Gaussian discriminant function. Motivated by neural collapse (Papyan et al., 2020), in which the class means align with the classifier weights, NCTTA (Chen et al., 2026) extends this alignment to individual samples and adapts the model so that each test feature moves toward the classifier weights of its likely classes. MixTTA (Jung et al., 2026) augments normalization layers with a low-rank cross-channel mixing branch, which makes the inter-channel correlation adaptable. Likewise, WCW adjusts the feature statistics to improve model robustness without any information about the source, focusing on the within-class covariance.

3 PRELIMINARIES

3.1 TOTAL COVARIANCE DECOMPOSITION BY CLASS CONDITIONING

For a penultimate-layer feature $z \in \mathbb{R}^D$ with class label $y \in \{1, \ldots, C\}$, the law of total covariance (Casella & Berger, 2002) decomposes the covariance of z into the expectation of the conditional covariance and the covariance of the conditional mean:

$$\mathrm{Cov}(z) = \mathbb{E}_{c\sim\pi}\big[\,\mathrm{Cov}(z \mid y = c)\big] + \mathrm{Cov}_{c\sim\pi}\left(\mathbb{E}[z \mid y = c]\right), \tag{1}$$

where the outer expectation and covariance are taken over the class label y, which follows the class prior π with $\pi_c = \Pr(y = c)$. In Eq. (1), the first term on the right-hand side is the class-averaged scatter of features about their own class mean, and the second is the scatter of the class means about the global mean. Accordingly, Eq. (1) can be rewritten as

$$\Sigma_{\mathrm{total}} = \Sigma_{\mathrm{within}} + \Sigma_{\mathrm{mean}}, \tag{2}$$

where Σ_{total} is the total covariance of z, Σ_{within} is its class-averaged within-class covariance, and Σ_{mean} is its class-mean covariance. This identity shows that the feature covariance can be exactly decomposed by conditioning on the class, and the total covariance alone does not reveal whether a deviation from the source stems from the arrangement of the class means or from the scatter within a class.

3.2 AFFINE TRANSFORMATION OF THE PENULTIMATE FEATURE

In this paper, we focus on affine transformations of the penultimate feature of the form

$$z' = T(z - m), \qquad T \in \mathbb{R}^{D\times D} \text{ invertible}, \quad m \in \mathbb{R}^D, \tag{3}$$

whose output is fed to the fixed linear head. We restrict our attention to this family for two reasons. First, the family can realize any adjustment of the first- and second-order moments of the features. Specifically, a translation by m can impose any target mean and an invertible linear map T can carry any covariance to any other. Second, the family contains the feature transformations of recent statistics-adjusting methods: NEO (Murphy et al., 2026) is a pure translation by the running mean, and TCA (You et al., 2025) transforms the feature covariance toward that of high-certainty test samples and shifts the mean accordingly. The activation shifting of FOA (Niu et al., 2024), which translates the class token toward the source mean, also belongs to the family. Hence, the analysis in Section 4 holds for each of these transformations under its own choice of T and m.

4 WITHIN-CLASS WHITENING

4.1 MITIGATION OF BOUNDARY CROSSING THROUGH WITHIN-CLASS WHITENING

Suppose that two classes i and j are separated by a decision boundary. Under domain shift, the covariance of each class is elongated, so a portion of the features can spill across the boundary and be misclassified. The following proposition bounds the probability of crossing the decision boundary.

Proposition 1. *Let n_{ij} denote the unit normal of the decision boundary between classes i and j, d_{ij} the distance from the mean of class i to the decision boundary, and σ_{ij}^2 the variance of class i along n_{ij}. If the mean of class i lies on its own side of the boundary, a feature of class i crosses the boundary with probability at most $\sigma_{ij}^2/(\sigma_{ij}^2 + d_{ij}^2)$.*

Since its proof in Appendix A.1 relies only on Cantelli's inequality, the bound holds for any feature distribution with finite variance. To reduce this bound for every pair of classes, d_{ij} should be increased and σ_{ij}^2 should be reduced. As Figure 2 shows, domain shift stretches the covariance of each class along certain directions, and existing TTA methods leave it distorted. Since the stretched covariance raises σ_{ij}^2 whenever n_{ij} lies along one of these directions, we concentrate on controlling this variance, and the following proposition shows how the affine transformation in Eq. (3) rescales it.

Proposition 2. *For a unit direction n, let $\sigma_y^2(n)$ denote the variance of the transformed features of class y along n, so that $\sigma_{ij}^2 = \sigma_i^2(n_{ij})$. Then, the class-averaged variance along the direction n is*

$$\mathbb{E}_{y\sim\pi}\left[\sigma_y^2(n)\right] = n^\top T \Sigma_{\text{within}} T^\top n, \tag{4}$$

and it equals one in every direction when $T\Sigma_{\text{within}}T^\top = I$.

Its proof is in Appendix A.2. As Eq. (4) describes, a transformation with $T\Sigma_{\text{within}}T^\top = I$ removes the influence of the elongated Σ_{within}, preventing σ_{ij}^2 from being inflated. However, this condition does not yield a unique solution: it holds for $T = Q\Sigma_{\text{within}}^{-1/2}$ with any orthogonal Q. The following proposition resolves this degree of freedom.

Proposition 3. *Let μ_c denote the feature mean of class c and $u = z - \mu_c$ the within-class residual. Among the transformations with $T\Sigma_{\text{within}}T^\top = I$, $T^* = \Sigma_{\text{within}}^{-1/2}$ is the unique one that minimizes $\mathbb{E}\left[\|Tu - u\|^2\right]$.*

The proposition follows from Kessy et al. (2018) and is proved in Appendix A.3. Combining the above propositions, we conclude that, among the operators that make the class-averaged within-class covariance of the transformed feature z' insensitive to the distortion of Σ_{within} under domain shift, the within-class whitening transformation $T^* = \Sigma_{\text{within}}^{-1/2}$ is the unique one that minimizes the displacement of the within-class residual.

4.2 Whitening the total covariance distorts the class prior

Restricting the whitening scope to Σ_{within} has a further advantage: whitening the total covariance implicitly interferes with the class prior π, whereas within-class whitening does not. To see how the prior is affected, we consider the transformation $T_\alpha = (\Sigma_{\text{within}} + \alpha\Sigma_{\text{mean}})^{-1/2}$ with $\alpha \geq 0$, which is equivalent to within-class whitening at $\alpha = 0$ and total whitening at $\alpha = 1$. The following proposition reveals the resulting impact on the class prior.

Proposition 4. *Let μ denote the global feature mean. Then,*

$$\left\|T_\alpha(\mu_c - \mu)\right\|^2 < \frac{1}{\alpha\,\pi_c}. \tag{5}$$

Its proof is in Appendix A.4. As α grows, the bound tightens and each transformed class mean is confined increasingly close to the global mean. The mean of a frequent class (high π_c) is drawn closer to the global mean and hence to the decision boundaries with adjacent classes, whereas that of a rare class (low π_c) stays farther away. Consequently, samples of frequent classes are more likely to be misclassified as rare classes; frequent classes are thus predicted less often than they occur, and rare classes more often. This flattens the distribution of predicted classes toward uniform, which amounts to an implicit injection of a uniform class prior. By contrast, whitening only the within-class covariance ($\alpha = 0$) is not subject to this bound and does not incur the uniform-prior injection of total whitening.

4.3 Online estimation of the whitening operator

As the analysis above suggests, $T^* = \Sigma_{\text{within}}^{-1/2}$ is the natural choice in the affine family of Eq. (3) to make the within-class covariance isotropic while suppressing the interference with the class prior. Building on T^*, we propose WCW, whose overall procedure is shown in Figure 3. Specifically, WCW accumulates the count and the first two moment sums of the arriving features. Every feature z of the batch updates the global count N and the moment sums s, S as

$$N \leftarrow N + 1, \quad s \leftarrow s + z, \quad S \leftarrow S + zz^\top, \tag{6}$$

and, for its predicted class $c = \hat{y}$ only, the class-wise count N_c and moment sum s_c as

$$N_c \leftarrow N_c + 1, \quad s_c \leftarrow s_c + z. \tag{7}$$

The global mean and the class-averaged within-class covariance then follow in closed form as

$$\hat{\mu} = \frac{s}{N}, \qquad \widehat{\Sigma}_{\text{within}} = \frac{S}{N} - \frac{1}{N}\sum_{c:N_c>0}\frac{s_c s_c^\top}{N_c}, \tag{8}$$

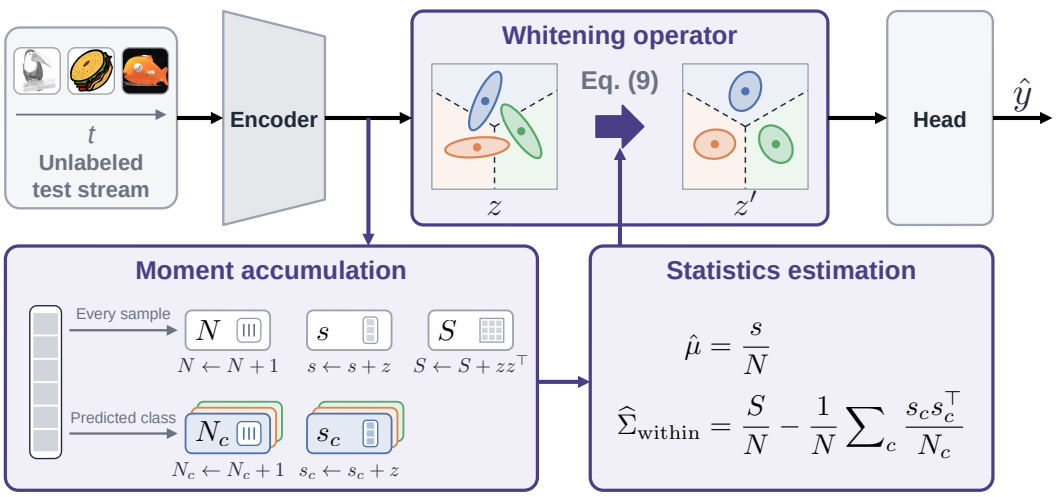


Figure 3: **Overview of WCW.** WCW accumulates the global and class-wise counts and moment sums of the features of each test batch, estimates the global mean $\hat{\mu}$ and the class-averaged within-class covariance $\widehat{\Sigma}_{\text{within}}$ from them, and transforms the penultimate feature.

which are derived in Appendix B. Given these estimates, we define the within-class whitening operator $\hat{T}$ and transform the feature z to z':

$$\hat{T} := (\widehat{\Sigma}_{\text{within}} + \varepsilon I)^{-1/2}, \qquad z' = \gamma\, \hat{T}(z - \hat{\mu}), \tag{9}$$

where γ is a coefficient that restores the scale of the feature norm, $\varepsilon > 0$ is a small constant that regularizes the inverse square root, and I denotes the identity matrix. The linear map $\hat{T}$ estimates T^* and whitens the within-class covariance of the features, while the translation by $\hat{\mu}$ corresponds to the re-centering strategy proposed by NEO (Murphy et al., 2026). Since $\hat{T}$ is built from the estimated within-class covariance alone, it requires neither source feature statistics nor target labels.

5 Experiments

5.1 Experimental protocol

Backbones, datasets, and baselines. We employ ViT-B/16 (Dosovitskiy et al., 2021) and ResNet-50 (He et al., 2016) with group normalization (Wu & He, 2018) as backbones. We evaluate on ImageNet-C (Hendrycks & Dietterich, 2019) at severity level 5, which applies 15 types of corruption to the validation images of ImageNet (Russakovsky et al., 2015), as well as on ImageNet-R (Hendrycks et al., 2021) and ImageNet-Sketch (Wang et al., 2019), whose renditions and sketches impose broader domain shifts. We adapt the model with three parameter-updating TTA methods, SAR (Niu et al., 2023), DeYO (Lee et al., 2024), and ReCAP (Hu et al., 2025), and combine each with three plug-ins, TCA (You et al., 2025), NEO (Murphy et al., 2026), and MixTTA (Jung et al., 2026), none of which requires any information about the source domain, so the nine combinations are compared with WCW in the same fully online setting. We also report results for the frozen backbone with WCW and with each plug-in except MixTTA, which requires parameter updates.

Scenarios and evaluation. On ImageNet-C, we follow the mild and three wild scenarios of Niu et al. (2023): an input stream with temporally correlated labels (imbalanced label shifts), adaptation on a single sample at a time (batch size 1), and a stream that mixes all 15 corruptions (mixed shifts). ImageNet-R and ImageNet-Sketch are evaluated in the mild scenario. All results are averaged over three seeds, with the standard deviation across seeds as a subscript. Further implementation details, including the hyperparameters of all employed methods, are reported in Appendix C.

Table 1: ResNet-50 GN top-1 accuracy (%). **Bold** values denote the top-performing results within each block. Since MixTTA requires parameter updates, the Source+MixTTA row is left blank.

Method	ImageNet-C				IN-R	IN-Sketch
	Mild	Label shifts	Batch size 1	Mixed shifts		
Source	$31.4_{\pm 0.0}$	$31.4_{\pm 0.0}$	$31.4_{\pm 0.0}$	$31.4_{\pm 0.0}$	$40.8_{\pm 0.0}$	$29.2_{\pm 0.0}$
+TCA	$32.5_{\pm 0.0}$	$32.8_{\pm 0.0}$	$32.5_{\pm 0.0}$	$33.7_{\pm 0.0}$	$\mathbf{42.5}_{\pm 0.0}$	$30.2_{\pm 0.0}$
+NEO	$34.3_{\pm 0.0}$	$34.2_{\pm 0.0}$	$34.3_{\pm 0.0}$	$32.1_{\pm 0.0}$	$41.8_{\pm 0.0}$	$30.5_{\pm 0.0}$
+MixTTA	—	—	—	—	—	—
+WCW	$\mathbf{36.4}_{\pm 0.0}$	$\mathbf{36.9}_{\pm 0.0}$	$\mathbf{36.4}_{\pm 0.0}$	$\mathbf{34.9}_{\pm 0.0}$	$42.1_{\pm 0.1}$	$\mathbf{32.4}_{\pm 0.0}$
SAR	$36.7_{\pm 1.3}$	$36.0_{\pm 1.2}$	$36.0_{\pm 0.0}$	$38.5_{\pm 0.1}$	$41.6_{\pm 0.0}$	$30.5_{\pm 0.1}$
+TCA	$38.5_{\pm 0.1}$	$38.5_{\pm 1.2}$	$37.2_{\pm 0.0}$	$39.6_{\pm 0.1}$	$43.2_{\pm 0.0}$	$31.6_{\pm 0.1}$
+NEO	$42.4_{\pm 0.0}$	$45.1_{\pm 0.1}$	$45.0_{\pm 0.0}$	$39.6_{\pm 0.0}$	$43.1_{\pm 0.0}$	$33.4_{\pm 0.1}$
+MixTTA	$37.5_{\pm 1.3}$	$37.5_{\pm 1.6}$	$36.7_{\pm 0.3}$	$38.3_{\pm 0.0}$	$41.7_{\pm 0.0}$	$31.5_{\pm 0.2}$
+WCW	$\mathbf{46.7}_{\pm 0.0}$	$\mathbf{49.7}_{\pm 0.1}$	$\mathbf{45.5}_{\pm 0.0}$	$\mathbf{42.3}_{\pm 0.0}$	$\mathbf{44.0}_{\pm 0.1}$	$\mathbf{36.4}_{\pm 0.1}$
DeYO	$45.2_{\pm 0.3}$	$41.3_{\pm 3.0}$	$45.8_{\pm 1.7}$	$36.6_{\pm 1.3}$	$44.2_{\pm 0.1}$	$34.5_{\pm 0.2}$
+TCA	$45.8_{\pm 0.2}$	$42.5_{\pm 2.3}$	$45.5_{\pm 0.8}$	$38.9_{\pm 1.4}$	$45.8_{\pm 0.0}$	$35.3_{\pm 0.2}$
+NEO	$49.8_{\pm 0.1}$	$51.1_{\pm 0.1}$	$\mathbf{52.4}_{\pm 0.1}$	$38.4_{\pm 1.4}$	$46.8_{\pm 0.1}$	$38.3_{\pm 0.2}$
+MixTTA	$44.4_{\pm 1.8}$	$42.1_{\pm 2.7}$	$45.8_{\pm 1.8}$	$40.1_{\pm 1.3}$	$44.8_{\pm 0.2}$	$36.0_{\pm 0.5}$
+WCW	$\mathbf{51.6}_{\pm 0.0}$	$\mathbf{52.7}_{\pm 0.0}$	$52.1_{\pm 0.1}$	$\mathbf{45.2}_{\pm 0.1}$	$\mathbf{47.8}_{\pm 0.1}$	$\mathbf{39.3}_{\pm 0.1}$
ReCAP	$46.4_{\pm 0.6}$	$42.6_{\pm 3.9}$	$47.5_{\pm 2.5}$	$42.1_{\pm 0.1}$	$45.3_{\pm 0.1}$	$36.0_{\pm 0.3}$
+TCA	$47.0_{\pm 0.6}$	$43.7_{\pm 3.8}$	$49.3_{\pm 0.6}$	$42.9_{\pm 0.1}$	$46.7_{\pm 0.1}$	$36.7_{\pm 0.2}$
+NEO	$50.4_{\pm 0.1}$	$51.5_{\pm 0.1}$	$\mathbf{53.3}_{\pm 0.0}$	$43.1_{\pm 0.1}$	$47.6_{\pm 0.1}$	$38.7_{\pm 0.2}$
+MixTTA	$48.0_{\pm 0.1}$	$43.5_{\pm 2.3}$	$47.3_{\pm 2.2}$	$43.2_{\pm 0.2}$	$45.8_{\pm 0.1}$	$36.8_{\pm 0.5}$
+WCW	$\mathbf{51.8}_{\pm 0.0}$	$\mathbf{52.5}_{\pm 0.1}$	$52.3_{\pm 0.0}$	$\mathbf{45.4}_{\pm 0.1}$	$\mathbf{48.0}_{\pm 0.1}$	$\mathbf{39.5}_{\pm 0.1}$

5.2 Main results

Table 1 presents the main results on ResNet-50 GN. WCW improves all parameter-updating methods regardless of the adaptation scenario or dataset, and records the highest accuracy among the plug-ins in most cases. The largest gains are 13.7 points for SAR under label shifts on ImageNet-C, 3.6 points for DeYO on ImageNet-R, and 5.9 points for SAR on ImageNet-Sketch. WCW also remains effective in the backpropagation-free setting: on the frozen backbone, it improves the mild-scenario accuracy from 31.4% to 36.4%, exceeding the gains of TCA and NEO. Beyond the consistent gain, WCW substantially stabilizes the adaptation procedure: the base methods show standard deviations of up to 3.9 points under label shifts and 2.5 points under batch size 1, whereas those of WCW never exceed 0.1 points in any setting. This stabilization is particularly pronounced for DeYO under mixed shifts, where the standard deviation of DeYO and its combinations with the other plug-ins ranges from 1.3 to 1.4 points, whereas WCW lowers it to only 0.1 points. On ViT-B/16, WCW records the highest accuracy among the plug-ins in 22 of the 24 cells, as Appendix E reports. Appendices D, F, and I report the computational overhead, the effect of pseudo-labels, and the per-corruption accuracy, respectively.

5.3 Ablation of the affine map

Table 2 ablates the two parts of the map in Eq. (9) in the mild scenario of ImageNet-C: the translation alone ($z' = z - \hat{\mu}$, which coincides with NEO), the whitening alone ($z' = \gamma \hat{T} z$), and the full map ($z' = \gamma \hat{T}(z - \hat{\mu})$). On ResNet-50 GN, the whitening alone matches the full map to within 0.1 points in every cell. On ViT-B/16, the two parts are complementary: the translation and the whitening contribute 0.6 and 1.0 points on average, respectively, whereas the full map contributes 1.5 points, and the whitening alone remains 0.4 to 0.5 points below the full map in every cell. Therefore, we conclude that the gain of WCW originates mainly from the whitening, and the mean subtraction contributes a consistent but minor gain only on ViT-B/16.

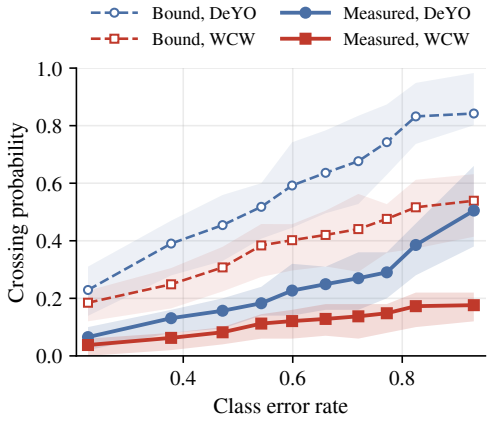


Figure 4: **Crossing probability against the class error rate of DeYO (ResNet-50 GN, zoom blur).** Classes are grouped into deciles of the class error rate; lines and bands show the decile means and interquartile ranges of the measured crossing probability (solid) and the bound of Proposition 1 (dashed).

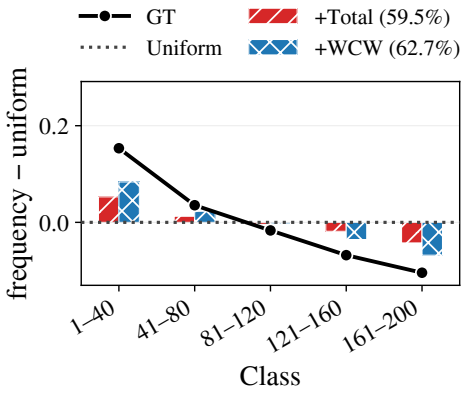


Figure 5: **Predicted class distribution on ImageNet-R (ViT-B/16, DeYO).** The line with markers denotes the ground-truth class frequency and the bars indicate the predicted frequency, both as deviations from the uniform distribution; the legend gives the accuracy of each whitening method.

Table 2: Ablation of the affine map.

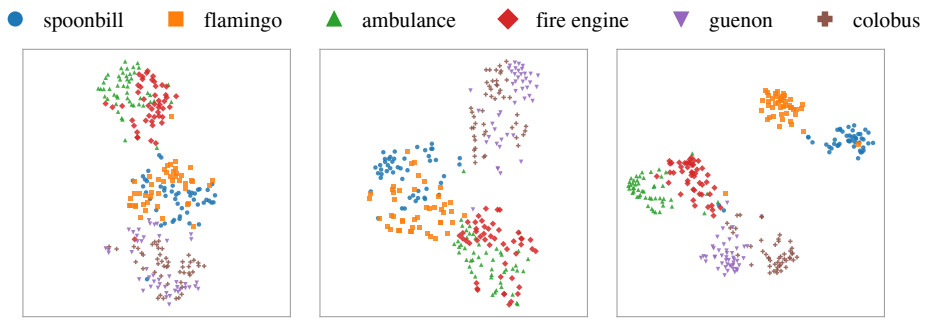
	Translation	Whitening	ViT-B/16		ResNet-50 GN	
			DeYO	ReCAP	DeYO	ReCAP
(1) Baseline			$64.3_{\pm 0.3}$	$64.0_{\pm 0.1}$	$45.2_{\pm 0.3}$	$46.4_{\pm 0.6}$
(2) NEO	✓		$65.0_{\pm 0.0}$	$64.5_{\pm 0.1}$	$49.8_{\pm 0.1}$	$50.4_{\pm 0.1}$
(3)		✓	$65.6_{\pm 0.0}$	$64.8_{\pm 0.0}$	$\mathbf{51.6}_{\pm 0.0}$	$51.7_{\pm 0.0}$
(4) WCW	✓	✓	$\mathbf{66.0}_{\pm 0.0}$	$\mathbf{65.3}_{\pm 0.0}$	$\mathbf{51.6}_{\pm 0.0}$	$\mathbf{51.8}_{\pm 0.0}$

5.4 Boundary-crossing probability and its bound

For ResNet-50 GN under the zoom blur corruption, Figure 4 groups the classes by their error rate under DeYO and reports the measured crossing probability of each group together with its upper bound from Proposition 1. Both the measured probability and its bound rise with the error rate, and WCW lowers both with a reduction that grows with the error rate. In the decile with the highest DeYO error rate, the crossing probability falls from 0.51 to 0.18 and the bound from 0.84 to 0.54. These observations support Propositions 1 and 2: the measured crossing probability never exceeds the bound, and the whitening operator of Proposition 3 lowers both the bound and the crossing probability.

5.5 Implicit uniform prior injection of total whitening

Section 4.2 argues that whitening with the total covariance flattens the predicted class distribution toward uniform. To verify this, we compare WCW with Total, which replaces the class-averaged within-class covariance in Eq. (9) with the total covariance, on ImageNet-R, whose class frequencies are imbalanced. Figure 5 groups the classes by their ground-truth label frequency and plots how far the predicted frequency of each group departs from the uniform distribution. Under Total, which remains at the accuracy of DeYO (59.5% vs. 59.6%), the predicted distribution moves toward uniform: frequent classes are predicted less often than they occur and rare classes more often. In contrast, WCW (62.7%) follows the true class proportions more closely, consistent with the analysis in Section 4.2. Appendix G sweeps α of T_α and shows that the accuracy decreases monotonically from WCW to Total.



(a) No adaptation (b) ReCAP (c) ReCAP + WCW

Figure 6: **t-SNE visualization of the penultimate features of three pairs of similar classes.**

Table 3: Comparison with the source oracle.

Baseline	Method	ViT-B/16			ResNet-50 GN		
		IN-C	IN-R	IN-Sketch	IN-C	IN-R	IN-Sketch
DeYO	WCW	$\mathbf{66.0}_{\pm 0.0}$	$\mathbf{62.7}_{\pm 0.1}$	$\mathbf{47.0}_{\pm 0.1}$	$\mathbf{51.6}_{\pm 0.0}$	$\mathbf{47.8}_{\pm 0.1}$	$\mathbf{39.3}_{\pm 0.1}$
	Oracle	$65.6_{\pm 0.0}$	$61.3_{\pm 0.0}$	$46.4_{\pm 0.1}$	$51.3_{\pm 0.0}$	$46.7_{\pm 0.2}$	$38.3_{\pm 0.1}$
ReCAP	WCW	$\mathbf{65.3}_{\pm 0.0}$	$\mathbf{61.8}_{\pm 0.1}$	$\mathbf{46.1}_{\pm 0.0}$	$\mathbf{51.8}_{\pm 0.0}$	$\mathbf{48.0}_{\pm 0.1}$	$\mathbf{39.5}_{\pm 0.1}$
	Oracle	$65.0_{\pm 0.0}$	$60.9_{\pm 0.1}$	$45.8_{\pm 0.1}$	$51.5_{\pm 0.0}$	$47.0_{\pm 0.2}$	$38.8_{\pm 0.2}$

5.6 t-SNE visualization of confusable-class features

To investigate how WCW deals with classes that resemble each other, we select three pairs of similar classes whose features lie in the same neighborhood under ReCAP—spoonbill and flamingo (birds), ambulance and fire engine (vehicles), and guenon and colobus (monkeys)—and plot the t-SNE embedding of their features in Figure 6 for ResNet-50 GN under the elastic transform corruption of ImageNet-C. Before and after adaptation with ReCAP, the two classes of each pair remain mixed in a single cluster. However, integrating WCW into ReCAP separates each class into its own cluster, indicating that WCW recovers the ability to discriminate between similar classes.

5.7 Is whitening enough? Oracle comparison with source recoloring

One may wonder whether realigning the within-class covariance with that of the source is better than whitening it. We therefore compare WCW with the source oracle, which recolors the class-averaged within-class covariance toward that of the source:

$$T_{\text{oracle}} = (\Sigma_{\text{within}}^{\text{source}} + \varepsilon I)^{1/2} \, (\widehat{\Sigma}_{\text{within}} + \varepsilon I)^{-1/2}, \tag{10}$$

where $\Sigma_{\text{within}}^{\text{source}}$ denotes the class-averaged within-class covariance of the source features. Table 3 reports the comparison between $\hat{T}$ and T_{oracle}. Although T_{oracle} restores the within-class covariance to its source counterpart, it underperforms WCW in all twelve settings. These results suggest that restoring the source structure brings no additional benefit and that whitening alone suffices.

6 Conclusion

We identified that existing TTA methods leave the within-class covariance largely uncontrolled, so that a portion of the samples can still be misclassified. Based on this finding, we proposed WCW, a plug-in module that whitens the penultimate features with the class-averaged within-class covariance estimated online from the predictions of the model. Our analysis shows that this whitening makes the features robust to the distortion of the within-class covariance and avoids the class-prior distortion caused by rearranging the class means. Requiring neither backpropagation nor pre-computation of source statistics, WCW improves existing parameter-updating methods and substantially stabilizes the adaptation procedure.

AI USE STATEMENT

In this work, we used generative AI tools for the following tasks with required disclosure: implementing the proposed method and the evaluation scripts on top of the published code of the baselines, executing the experiments and collecting their logs, and assisting in the writing of the proofs of Propositions 1 to 4. We have not used generative AI tools to develop the theoretical model or conceptual framework, formulate the mathematical claims (the statements of Propositions 1 to 4), provide critical ingredients for the proofs, propose or refine hypotheses, design the methodology and experiments, interpret the results, generate synthetic data, translate text, or clean and reformat datasets; qualitative and thematic data analysis is not applicable to this work. Additionally, we used generative AI tools for tasks with recommended disclosure: drafting and editing parts of the paper to improve readability, creating and revising the figures, identifying related literature, and formatting references. We have reviewed all AI-assisted work and take responsibility for the final content of this work, including text, claims, and artifacts produced with the aid of generative AI.

ETHICS STATEMENT

This work studies test-time adaptation of image classifiers on the public benchmarks ImageNet-C, ImageNet-R, and ImageNet-Sketch with pretrained models released by their authors, used in accordance with their terms of use; it involves no human subjects, no personal data, and no newly collected data. The proposed method adapts a model with unlabeled test data at deployment, so its predictions inherit the biases of the pretrained model and of the test stream, and the pseudo-labels it relies on can reinforce existing errors under severe distribution shift. The reported accuracy therefore provides no evidence of fairness across groups or of safe behavior in a specific application, and we recommend application-specific validation and monitoring before deployment. We do not foresee other ethical concerns beyond those common to adaptation methods for image classification.

REPRODUCIBILITY STATEMENT

Section 4.3 and Figure 3 specify the proposed method, and Appendix B derives its closed-form estimator from the accumulated statistics. The assumptions and the complete proofs of Propositions 1 to 4 are given in Appendix A. Section 5.1 describes the backbones, datasets, baselines, and evaluation scenarios, and Appendix C lists every hyperparameter in Table 4 together with the warm-up and refresh rules of the operator and the settings inherited from the published implementations of the baselines. All reported accuracies are averages over three seeds with their standard deviations, the per-corruption results are reported in Appendix I, and the protocol of the cost measurement is stated in Appendix D. The source code will be released upon acceptance.

REFERENCES

George Casella and Roger L. Berger. *Statistical Inference*. Duxbury, Pacific Grove, CA, 2nd edition, 2002.

Dian Chen, Dequan Wang, Trevor Darrell, and Sayna Ebrahimi. Contrastive test-time adaptation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 295–305, 2022. URL https://openaccess.thecvf.com/content/CVPR2022/html/Chen_Contrastive_Test-Time_Adaptation_CVPR_2022_paper.html.

Xiao Chen, Zhongjing Du, Jiazhen Huang, Xu Jiang, Li Lu, Jingyan Jiang, and Zhi Wang. Neural collapse in test-time adaptation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 10567–10576, 2026.

Mario Döbler, Robert A. Marsden, and Bin Yang. Robust mean teacher for continual and gradual test-time adaptation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 7704–7714, 2023. URL https://openaccess.thecvf.com/content/CVPR2023/html/Dobler_Robust_Mean_Teacher_for_Continual_and_Gradual_Test-Time_Adaptation_CVPR_2023_paper.html.

Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszko-

reit, and Neil Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale. In *International Conference on Learning Representations*, 2021. URL https://dblp.org/rec/conf/iclr/DosovitskiyB0WZ21.

Yossi Gandelsman, Yu Sun, Xinlei Chen, and Alexei A. Efros. Test-time training with masked autoencoders. In *Advances in Neural Information Processing Systems*, 2022. URL https://proceedings.neurips.cc/paper_files/paper/2022/hash/bcdec1c2d60f94a93b6e36f937aa0530-Abstract-Conference.html.

Augustine H. Gray and John D. Markel. A spectral-flatness measure for studying the autocorrelation method of linear prediction of speech analysis. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 22(3):207–217, 1974.

Zongbo Han, Jialong Yang, Guangyu Wang, Junfan Li, Qianli Xu, Mike Zheng Shou, and Changqing Zhang. DOTA: Distributional test-time adaptation of vision-language models. In *Advances in Neural Information Processing Systems*, 2025.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770–778, 2016. URL https://dblp.org/rec/conf/cvpr/HeZRS16.

Dan Hendrycks and Thomas G. Dietterich. Benchmarking neural network robustness to common corruptions and perturbations. In *International Conference on Learning Representations*, 2019. URL https://dblp.org/rec/conf/iclr/HendrycksD19.

Dan Hendrycks, Steven Basart, Norman Mu, Saurav Kadavath, Frank Wang, Evan Dorundo, Rahul Desai, Tyler Zhu, Samyak Parajuli, Mike Guo, Dawn Song, Jacob Steinhardt, and Justin Gilmer. The many faces of robustness: A critical analysis of out-of-distribution generalization. In *IEEE International Conference on Computer Vision*, pp. 8320–8329, 2021. URL https://dblp.org/rec/conf/iccv/HendrycksBMKWDD21.

Zixuan Hu, Yichun Hu, Xiaotong Li, Shixiang Tang, and Ling-Yu Duan. Beyond entropy: Region confidence proxy for wild test-time adaptation. In *International Conference on Machine Learning*, 2025. URL https://dblp.org/rec/conf/icml/HuHLTD25.

Yusuke Iwasawa and Yutaka Matsuo. Test-time classifier adjustment module for model-agnostic domain generalization. In *Advances in Neural Information Processing Systems*, 2021. URL https://proceedings.neurips.cc/paper/2021/hash/1415fe9fea0fa1e45dddcff5682239a0-Abstract.html.

Mansoo Jung, Youngwook Kim, and Jungwoo Lee. MixTTA: Low-rank cross-channel mixing for reliable test-time adaptation. In *European Conference on Computer Vision*, 2026. URL https://doi.org/10.48550/arXiv.2606.28142. To appear.

Sanghun Jung, Jungsoo Lee, Nanhee Kim, Amirreza Shaban, Byron Boots, and Jaegul Choo. CAFA: Class-aware feature alignment for test-time adaptation. In *IEEE/CVF International Conference on Computer Vision*, pp. 19060–19071, 2023. URL https://openaccess.thecvf.com/content/ICCV2023/html/Jung_CAFA_Class-Aware_Feature_Alignment_for_Test-Time_Adaptation_ICCV_2023_paper.html.

Agnan Kessy, Alex Lewin, and Korbinian Strimmer. Optimal whitening and decorrelation. *The American Statistician*, 72(4):309–314, 2018.

Jonghyun Lee, Dahuin Jung, Saehyung Lee, Junsung Park, Juhyeon Shin, Uiwon Hwang, and Sungroh Yoon. Entropy is not enough for test-time adaptation: From the perspective of disentangled factors. In *International Conference on Learning Representations*, 2024. URL https://dblp.org/rec/conf/iclr/0004JLPSHY24.

Yanghao Li, Naiyan Wang, Jianping Shi, Xiaodi Hou, and Jiaying Liu. Adaptive batch normalization for practical domain adaptation. *Pattern Recognition*, 80:109–117, 2018.

Xiao Ma, Young D. Kwon, Pan Zhou, and Dong Ma. Architecture-agnostic test-time adaptation via backprop-free embedding alignment. In *International Conference on Learning Representations*, 2026.

Muhammad Jehanzeb Mirza, Pol Jané Soneira, Wei Lin, Mateusz Kozinski, Horst Possegger, and Horst Bischof. ActMAD: Activation matching to align distributions for test-time-training. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 24152–24161, 2023.

Alexander Murphy, Michal Danilowski, Soumyajit Chatterjee, and Abhirup Ghosh. NEO: No-optimization test-time adaptation through latent re-centering. In *International Conference on Learning Representations*, 2026.

A. Tuan Nguyen, Thanh Nguyen-Tang, Ser-Nam Lim, and Philip H. S. Torr. TIPI: Test time adaptation with transformation invariance. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 24162–24171, 2023. URL https://openaccess.thecvf.com/content/CVPR2023/html/Nguyen_TIPI_Test_Time_Adaptation_With_Transformation_Invariance_CVPR_2023_paper.html.

Shuaicheng Niu, Jiaxiang Wu, Yifan Zhang, Yaofo Chen, Shijian Zheng, Peilin Zhao, and Mingkui Tan. Efficient test-time model adaptation without forgetting. In *International Conference on Machine Learning*, pp. 16888–16905, 2022. URL https://dblp.org/rec/conf/icml/NiuW0CZZT22.

Shuaicheng Niu, Jiaxiang Wu, Yifan Zhang, Zhiquan Wen, Yaofo Chen, Peilin Zhao, and Mingkui Tan. Towards stable test-time adaptation in dynamic wild world. In *International Conference on Learning Representations*, 2023. URL https://dblp.org/rec/conf/iclr/Niu00WCZT23.

Shuaicheng Niu, Chunyan Miao, Guohao Chen, Pengcheng Wu, and Peilin Zhao. Test-time model adaptation with only forward passes. In *International Conference on Machine Learning*, 2024.

Vardan Papyan, X. Y. Han, and David L. Donoho. Prevalence of neural collapse during the terminal phase of deep learning training. *Proceedings of the National Academy of Sciences*, 117(40): 24652–24663, 2020. doi: 10.1073/pnas.2015509117.

Subhankar Roy, Aliaksandr Siarohin, Enver Sangineto, Samuel Rota Bulò, Nicu Sebe, and Elisa Ricci. Unsupervised domain adaptation using feature-whitening and consensus loss. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 9463–9472, 2019. doi: 10.1109/CVPR.2019.00970.

Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael S. Bernstein, Alexander C. Berg, and Li Fei-Fei. ImageNet large scale visual recognition challenge. *International Journal of Computer Vision*, 115(3):211–252, 2015.

Steffen Schneider, Evgenia Rusak, Luisa Eck, Oliver Bringmann, Wieland Brendel, and Matthias Bethge. Improving robustness against common corruptions by covariate shift adaptation. In *Advances in Neural Information Processing Systems*, 2020. URL https://dblp.org/rec/conf/nips/SchneiderRE0BB20.

Yongyi Su, Xun Xu, and Kui Jia. Revisiting realistic test-time training: Sequential inference and adaptation by anchored clustering. In *Advances in Neural Information Processing Systems*, 2022. URL https://dblp.org/rec/conf/nips/SuXJ22.

Baochen Sun, Jiashi Feng, and Kate Saenko. Return of frustratingly easy domain adaptation. In *AAAI Conference on Artificial Intelligence*, pp. 2058–2065, 2016. doi: 10.1609/AAAI.V30I1.10306. URL https://dblp.org/rec/conf/aaai/SunFS16.

Yu Sun, Xiaolong Wang, Zhuang Liu, John Miller, Alexei A. Efros, and Moritz Hardt. Test-time training with self-supervision for generalization under distribution shifts. In *International Conference on Machine Learning*, pp. 9229–9248, 2020. URL https://dblp.org/rec/conf/icml/SunWLMEH20.

Dequan Wang, Evan Shelhamer, Shaoteng Liu, Bruno A. Olshausen, and Trevor Darrell. Tent: Fully test-time adaptation by entropy minimization. In *International Conference on Learning Representations*, 2021. URL https://dblp.org/rec/conf/iclr/WangSLOD21.

Haohan Wang, Songwei Ge, Zachary C. Lipton, and Eric P. Xing. Learning robust global representations by penalizing local predictive power. In *Advances in Neural Information Processing Systems*, pp. 10506–10518, 2019. URL https://dblp.org/rec/conf/nips/WangGLX19.

Qin Wang, Olga Fink, Luc Van Gool, and Dengxin Dai. Continual test-time domain adaptation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 7201–7211, 2022. URL https://openaccess.thecvf.com/content/CVPR2022/html/Wang_Continual_Test-Time_Domain_Adaptation_CVPR_2022_paper.html.

Shuai Wang, Daoan Zhang, Zipei Yan, Jianguo Zhang, and Rui Li. Feature alignment and uniformity for test time adaptation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 20050–20060, 2023. URL https://openaccess.thecvf.com/content/CVPR2023/html/Wang_Feature_Alignment_and_Uniformity_for_Test_Time_Adaptation_CVPR_2023_paper.html.

Yuxin Wu and Kaiming He. Group normalization. In *European Conference on Computer Vision*, pp. 3–19, 2018. doi: 10.1007/978-3-030-01261-8_1. URL https://dblp.org/rec/conf/eccv/WuH18.

Linjing You, Jiabao Lu, and Xiayuan Huang. Test-time correlation alignment. In *International Conference on Machine Learning*, 2025. URL https://dblp.org/rec/conf/icml/YouLH25.

Marvin Zhang, Sergey Levine, and Chelsea Finn. MEMO: Test time robustness via adaptation and augmentation. In *Advances in Neural Information Processing Systems*, 2022. URL https://proceedings.neurips.cc/paper_files/paper/2022/hash/fc28053a08f59fccb48b11f2e31e81c7-Abstract-Conference.html.

Qingyang Zhang, Yatao Bian, Xinke Kong, Peilin Zhao, and Changqing Zhang. COME: Test-time adaption by conservatively minimizing entropy. In *International Conference on Learning Representations*, 2025a.

Youjia Zhang, Youngeun Kim, Young-Geun Choi, Hongyeob Kim, Huiling Liu, and Sungeun Hong. Backpropagation-free test-time adaptation via probabilistic Gaussian alignment. In *Advances in Neural Information Processing Systems*, 2025b.

A PROOFS OF PROPOSITIONS 1–4

Throughout, we assume $\Sigma_{\text{within}} \succ 0$, and $\Sigma^{-1/2}$ denotes the unique symmetric inverse square root of a symmetric positive definite matrix Σ.

A.1 PROOF OF PROPOSITION 1

Proof. Let w_i and b_i denote the weight vector and the bias of the head for class i. The decision boundary between classes i and j is the set of features z with $(w_i - w_j)^\top z + b_i - b_j = 0$, and its unit normal is $n_{ij} = (w_i - w_j)/\|w_i - w_j\|$. The signed distance from a feature z to the boundary is

$$g(z) = \frac{(w_i - w_j)^\top z + b_i - b_j}{\|w_i - w_j\|}, \tag{11}$$

which is positive on the side where class i lies, so that a feature of class i crosses the boundary exactly when $g(z) < 0$. Since g is affine in z, its mean over the features of class i is the signed distance of the class mean, which equals d_{ij} when the mean of class i lies on that side, and its variance is

$$\operatorname{Var}\big(g(z) \mid y = i\big) = n_{ij}^\top \operatorname{Cov}(z \mid y = i)\, n_{ij} = \sigma_{ij}^2. \tag{12}$$

Cantelli's inequality then gives

$$\Pr\big(g(z) < 0 \mid y = i\big) \le \Pr\big(g(z) - d_{ij} \le -d_{ij} \mid y = i\big) \le \frac{\sigma_{ij}^2}{\sigma_{ij}^2 + d_{ij}^2}. \tag{13}$$

□

A.2 PROOF OF PROPOSITION 2

Proof. For the transformed feature $z' = T(z - m)$ of Eq. (3) and a unit direction n, the variance of class y along n is

$$\sigma_y^2(n) = \operatorname{Var}\big(n^\top z' \mid y\big) = n^\top T \operatorname{Cov}(z \mid y)\, T^\top n. \tag{14}$$

Averaging over $y \sim \pi$ and using $\mathbb{E}_{y\sim\pi}[\operatorname{Cov}(z \mid y)] = \Sigma_{\text{within}}$ gives Eq. (4). If $T\Sigma_{\text{within}}T^\top = I$, the right-hand side of Eq. (4) equals $n^\top n = 1$ for every unit direction n. The condition $T\Sigma_{\text{within}}T^\top = I$ states that $T\Sigma_{\text{within}}^{1/2}$ is orthogonal, so its solutions are exactly $T = Q\Sigma_{\text{within}}^{-1/2}$ with Q orthogonal. □

A.3 PROOF OF PROPOSITION 3

Proof. Let $T = Q\Sigma_{\text{within}}^{-1/2}$ with Q orthogonal. Since $\mathbb{E}[uu^\top] = \mathbb{E}_{c\sim\pi}[\operatorname{Cov}(z \mid y = c)] = \Sigma_{\text{within}}$,

$$\begin{aligned} \mathbb{E}\big[\|Tu - u\|^2\big] &= \operatorname{tr}(T\Sigma_{\text{within}}T^\top) - 2\operatorname{tr}(T\Sigma_{\text{within}}) + \operatorname{tr}(\Sigma_{\text{within}}) \\ &= D - 2\operatorname{tr}\big(Q\Sigma_{\text{within}}^{1/2}\big) + \operatorname{tr}(\Sigma_{\text{within}}), \end{aligned} \tag{15}$$

so the displacement is minimized when $\operatorname{tr}(Q\Sigma_{\text{within}}^{1/2})$ is maximized. Write $\Sigma_{\text{within}} = U\Lambda U^\top$ with $\Lambda = \operatorname{diag}(\lambda_1, \dots, \lambda_D)$, $\lambda_i > 0$, and let $G = U^\top QU$, which is orthogonal. Then

$$\operatorname{tr}(Q\Sigma_{\text{within}}^{1/2}) = \operatorname{tr}(G\Lambda^{1/2}) = \sum_{i=1}^{D} \lambda_i^{1/2}\, G_{ii}. \tag{16}$$

The rows of an orthogonal matrix have unit norm, so $G_{ii} \le 1$ for every i, with equality for every i only if $G = I$. Hence the trace is maximized uniquely at $G = I$, that is, at $Q = I$, which is Proposition 1 of Kessy et al. (2018). □

A.4 PROOF OF PROPOSITION 4

Proof. Let $v_c = \mu_c - \mu$ and $\tilde{a}_c = \Sigma_{\text{within}}^{-1/2} v_c$ for every class c, so that $\Sigma_{\text{mean}} = \sum_k \pi_k v_k v_k^\top$ and

$$B := \Sigma_{\text{within}}^{-1/2}\Sigma_{\text{mean}}\Sigma_{\text{within}}^{-1/2} = \sum_k \pi_k \tilde{a}_k \tilde{a}_k^\top. \tag{17}$$

Since $\Sigma_{\text{within}} + \alpha\Sigma_{\text{mean}} = \Sigma_{\text{within}}^{1/2}(I + \alpha B)\Sigma_{\text{within}}^{1/2}$,

$$\left\|T_\alpha v_c\right\|^2 = v_c^\top(\Sigma_{\text{within}} + \alpha\Sigma_{\text{mean}})^{-1}v_c = \tilde{a}_c^\top(I + \alpha B)^{-1}\tilde{a}_c. \tag{18}$$

Separate the contribution of class c from the other classes by writing $I + \alpha B = K^{-1} + \alpha\pi_c\tilde{a}_c\tilde{a}_c^\top$ with

$$K = \Big(I + \alpha\sum_{k\neq c}\pi_k\tilde{a}_k\tilde{a}_k^\top\Big)^{-1}, \qquad \kappa_c = \tilde{a}_c^\top K\tilde{a}_c \geq 0. \tag{19}$$

The Sherman–Morrison formula gives

$$\tilde{a}_c^\top(I + \alpha B)^{-1}\tilde{a}_c = \kappa_c - \frac{\alpha\pi_c\kappa_c^2}{1 + \alpha\pi_c\kappa_c} = \frac{\kappa_c}{1 + \alpha\pi_c\kappa_c} < \frac{1}{\alpha\pi_c}, \tag{20}$$

where the last inequality holds for every $\alpha > 0$ because $\alpha\pi_c\kappa_c < 1 + \alpha\pi_c\kappa_c$. This is Eq. (5). □

B Derivation of the within-class covariance estimator

By the definition in Eq. (1),

$$\Sigma_{\text{within}} = \mathbb{E}_{c\sim\pi}\big[\operatorname{Cov}(z \mid y = c)\big] = \sum_{c=1}^{C}\pi_c \operatorname{Cov}(z \mid y = c). \tag{21}$$

We estimate it from the N accumulated features by replacing the label with the predicted class $\hat{y}_i$, the prior π_c with the empirical frequency N_c/N, and the covariance of class c with the sample covariance about the class mean $\hat{\mu}_c = s_c/N_c$:

$$\begin{aligned}\widehat{\Sigma}_{\text{within}} &= \sum_{c:N_c>0}\frac{N_c}{N}\cdot\frac{1}{N_c}\sum_{i:\hat{y}_i=c}(z_i - \hat{\mu}_c)(z_i - \hat{\mu}_c)^\top \\ &= \frac{1}{N}\sum_{c:N_c>0}\sum_{i:\hat{y}_i=c}(z_i - \hat{\mu}_c)(z_i - \hat{\mu}_c)^\top.\end{aligned} \tag{22}$$

Expanding the outer products with $s_c = \sum_{i:\hat{y}_i=c} z_i = N_c\hat{\mu}_c$ gives, for each class,

$$\begin{aligned}\sum_{i:\hat{y}_i=c}(z_i - \hat{\mu}_c)(z_i - \hat{\mu}_c)^\top &= \sum_{i:\hat{y}_i=c} z_i z_i^\top - s_c\hat{\mu}_c^\top - \hat{\mu}_c s_c^\top + N_c\hat{\mu}_c\hat{\mu}_c^\top \\ &= \sum_{i:\hat{y}_i=c} z_i z_i^\top - \frac{s_c s_c^\top}{N_c}.\end{aligned} \tag{23}$$

Substituting this into Eq. (22) and using $\sum_c\sum_{i:\hat{y}_i=c} z_i z_i^\top = S$ yields Eq. (8).

C Implementation details

Since $\widehat{\Sigma}_{\text{within}}$ can be rank deficient early in the stream, the operator $\hat{T}$ is kept as the identity until both 24 batches have been processed and the number of accumulated samples N reaches twice the feature dimension D. The operator $\hat{T}$ is recomputed from the accumulated sums every 10 batches by a symmetric eigendecomposition. The regularizer ε is set to 0.01 times the mean eigenvalue of $\widehat{\Sigma}_{\text{within}}$ so that it scales with the features. The factor γ is the mean feature norm of the batch before the transformation divided by that after it, so that the transformation changes the shape of the features but not their overall scale. All other hyperparameters are inherited from the published settings of each method and are not tuned for WCW. Table 4 lists them.

Table 4: Hyperparameters. C is the number of classes and D the feature dimension (768 for ViT-B/16 and 2048 for ResNet-50 GN); values that differ between the backbones are given as ViT / ResNet.

Setting	Value
General	
Optimizer	SGD with momentum 0.9; SAM wrapper with $\rho = 0.05$ for SAR
Batch size	64; 1 in the batch-size-1 scenario
Learning rate	1×10^{-3} / 2.5×10^{-4}
Learning rate at batch size 1	3.1×10^{-5} / 1.6×10^{-5}
SAR	
Entropy threshold E_0	$0.4 \ln C$
Recovery threshold e_m	0.2
DeYO	
Entropy threshold	$0.5 \ln C$
Reweighting margin Ent_0	$0.4 \ln C$
PLPD threshold	0.2, with 4×4 patch shuffling
ReCAP	
Regional-entropy threshold τ_{RE}	$1.0 \ln C$ / $0.8 \ln C$
Reference entropy L_0	$1.0 \ln C$ / $0.8 \ln C$
Reweighting threshold	1.5 / 3.0
TCA	
Supports per pseudo-class K	100
CORAL iterations and step size	20 and 10^{-3}
MixTTA	
Rank r	4
Spectral projection strength η	0.9
WCW	
Warm-up	identity for the first 24 batches and while $N < 2D$
Refresh of $\hat{T}$	every 10 batches
Regularizer ε	0.01 times the mean eigenvalue of $\widehat{\Sigma}_{\mathrm{within}}$
Scale γ	mean feature norm of the batch before the transformation over that after it

D COMPUTATIONAL COST

Table 5 reports the throughput and the peak GPU memory of SAR, DeYO, and ReCAP with and without WCW on both backbones. Each configuration is run three times over the 50,000 images of the Gaussian noise corruption of ImageNet-C at severity 5 in the mild scenario, with batch size 64 and eight data-loading workers, on a single, otherwise idle NVIDIA RTX A6000. The throughput is the number of images per second over the whole stream, including data loading, averaged over the three runs. The memory is the peak usage of the process reported by `nvidia-smi` at intervals of 0.5 seconds, which includes the CUDA context. As reported in Table 5, WCW lowers the throughput by less than 1% on ViT-B/16 and by 7 to 10% on ResNet-50 GN, and raises the peak memory by 0.12 to 0.19 GiB, amounting to a modest overhead.

Table 5: Throughput and peak GPU memory with and without WCW on an NVIDIA RTX A6000 (Gaussian noise corruption of ImageNet-C). Throughput is the number of images per second over the whole stream and memory is the peak usage of the process in GiB.

		Throughput (images/s)		Memory (GiB)	
Backbone	Method	Baseline	+WCW	Baseline	+WCW
ViT-B/16	SAR	97.9	97.7 (−0.2%)	7.04	7.16 (+1.7%)
	DeYO	130.7	129.7 (−0.8%)	7.44	7.57 (+1.7%)
	ReCAP	130.1	128.9 (−0.9%)	7.37	7.50 (+1.8%)
ResNet-50 GN	SAR	238.4	222.1 (−6.8%)	6.61	6.80 (+2.9%)
	DeYO	331.7	298.8 (−9.9%)	6.72	6.91 (+2.8%)
	ReCAP	324.3	293.8 (−9.4%)	6.98	7.14 (+2.3%)

E MAIN RESULTS ON VIT-B/16

Table 6 reports the ViT-B/16 counterpart of Table 1. WCW improves every parameter-updating method in all six settings, and the largest gain on each dataset occurs on SAR: 4.4 points under label shifts on ImageNet-C, 8.4 points on ImageNet-R, and 15.4 points on ImageNet-Sketch. It also stabilizes the adaptation: the standard deviation across the three seeds reaches up to 1.0 point for the base methods, whereas it never exceeds 0.1 points with WCW.

Table 6: ViT-B/16 top-1 accuracy (%). The format follows Table 1.

Method	ImageNet-C				IN-R	IN-Sketch
	Mild	Label shifts	Batch size 1	Mixed shifts		
Source	$31.6_{\pm 0.0}$	$31.6_{\pm 0.0}$	$31.6_{\pm 0.0}$	$31.6_{\pm 0.0}$	$43.2_{\pm 0.0}$	$18.2_{\pm 0.0}$
+TCA	$32.1_{\pm 0.0}$	$32.4_{\pm 0.0}$	$32.1_{\pm 0.0}$	$32.5_{\pm 0.0}$	$45.4_{\pm 0.0}$	$18.9_{\pm 0.0}$
+NEO	$35.0_{\pm 0.0}$	$34.8_{\pm 0.0}$	$35.0_{\pm 0.0}$	$33.5_{\pm 0.0}$	$45.1_{\pm 0.0}$	$20.8_{\pm 0.0}$
+MixTTA	—	—	—	—	—	—
+WCW	$\mathbf{37.1}_{\pm 0.0}$	$\mathbf{37.4}_{\pm 0.1}$	$\mathbf{37.1}_{\pm 0.0}$	$\mathbf{35.7}_{\pm 0.0}$	$\mathbf{46.8}_{\pm 0.0}$	$\mathbf{22.7}_{\pm 0.0}$
SAR	$56.8_{\pm 0.2}$	$59.0_{\pm 0.5}$	$57.8_{\pm 0.2}$	$57.7_{\pm 0.0}$	$47.4_{\pm 0.4}$	$25.2_{\pm 0.6}$
+TCA	$57.0_{\pm 0.1}$	$59.9_{\pm 0.3}$	$58.0_{\pm 0.2}$	$58.1_{\pm 0.0}$	$49.9_{\pm 0.6}$	$29.2_{\pm 1.0}$
+NEO	$58.4_{\pm 0.0}$	$61.2_{\pm 0.1}$	$59.9_{\pm 0.1}$	$58.0_{\pm 0.0}$	$53.4_{\pm 0.1}$	$38.2_{\pm 0.1}$
+MixTTA	$57.2_{\pm 0.2}$	$60.8_{\pm 0.2}$	$57.9_{\pm 0.2}$	$58.9_{\pm 0.1}$	$49.6_{\pm 0.3}$	$35.6_{\pm 0.1}$
+WCW	$\mathbf{60.7}_{\pm 0.0}$	$\mathbf{63.4}_{\pm 0.0}$	$\mathbf{61.1}_{\pm 0.1}$	$\mathbf{59.5}_{\pm 0.0}$	$\mathbf{55.8}_{\pm 0.1}$	$\mathbf{40.6}_{\pm 0.0}$
DeYO	$64.3_{\pm 0.3}$	$64.8_{\pm 0.0}$	$64.3_{\pm 0.2}$	$59.6_{\pm 0.1}$	$59.6_{\pm 0.1}$	$43.3_{\pm 0.3}$
+TCA	$64.5_{\pm 0.2}$	$65.0_{\pm 0.1}$	$64.5_{\pm 0.4}$	$60.0_{\pm 0.1}$	$61.3_{\pm 0.1}$	$43.6_{\pm 0.3}$
+NEO	$65.0_{\pm 0.0}$	$65.0_{\pm 0.0}$	$65.7_{\pm 0.0}$	$59.9_{\pm 0.1}$	$61.4_{\pm 0.1}$	$44.8_{\pm 0.1}$
+MixTTA	$65.6_{\pm 0.1}$	$65.6_{\pm 0.1}$	$66.2_{\pm 0.1}$	$\mathbf{62.5}_{\pm 0.1}$	$58.2_{\pm 0.3}$	$43.7_{\pm 0.1}$
+WCW	$\mathbf{66.0}_{\pm 0.0}$	$\mathbf{66.3}_{\pm 0.0}$	$\mathbf{66.4}_{\pm 0.0}$	$62.0_{\pm 0.0}$	$\mathbf{62.7}_{\pm 0.1}$	$\mathbf{47.0}_{\pm 0.1}$
ReCAP	$64.0_{\pm 0.1}$	$64.7_{\pm 0.1}$	$64.8_{\pm 1.0}$	$60.0_{\pm 0.1}$	$60.1_{\pm 0.1}$	$43.3_{\pm 0.1}$
+TCA	$64.1_{\pm 0.0}$	$63.9_{\pm 1.5}$	$65.6_{\pm 0.4}$	$60.3_{\pm 0.1}$	$61.5_{\pm 0.1}$	$43.6_{\pm 0.1}$
+NEO	$64.5_{\pm 0.1}$	$64.9_{\pm 0.1}$	$66.1_{\pm 0.3}$	$60.2_{\pm 0.1}$	$61.0_{\pm 0.1}$	$44.6_{\pm 0.1}$
+MixTTA	$64.9_{\pm 0.1}$	$65.7_{\pm 0.1}$	$66.8_{\pm 0.1}$	$\mathbf{62.4}_{\pm 0.1}$	$58.2_{\pm 0.5}$	$42.9_{\pm 0.0}$
+WCW	$\mathbf{65.3}_{\pm 0.0}$	$\mathbf{66.0}_{\pm 0.1}$	$\mathbf{66.9}_{\pm 0.0}$	$61.8_{\pm 0.0}$	$\mathbf{61.8}_{\pm 0.1}$	$\mathbf{46.1}_{\pm 0.0}$

F PSEUDO-LABELS VERSUS GROUND-TRUTH LABELS IN THE WITHIN-CLASS STATISTICS

WCW estimates the class-averaged within-class covariance from the predictions of the model. To examine the influence of approximating the true covariance Σ_{within} with its estimate $\widehat{\Sigma}_{\text{within}}$, we

replace the predicted labels in the class sums of Eq. (8) with the ground-truth labels of each batch. Table 7 compares the two in the mild scenario of ImageNet-C: the pseudo-labels retain 81 to 96% of the gain obtained with ground-truth labels over the base methods. Notably, even for the frozen backbone, which starts from 31.6% on ViT-B/16 and 31.4% on ResNet-50 GN, the pseudo-labels retain 93% and 91% of the gain obtained with the ground-truth labels. This suggests that whitening with $\widehat{\Sigma}_{\text{within}}$ performs comparably to whitening with Σ_{within}.

Table 7: Comparison of whitening based on pseudo-labels and on ground-truth labels. Retained denotes the ratio of the gain of WCW with pseudo-labels over the base method to the gain with ground-truth labels; Source denotes the frozen backbone without adaptation.

			+WCW		
Backbone	Method	Base	Pseudo-labels	Ground truth	Retained
ViT-B/16	Source	$31.6_{\pm 0.0}$	$37.1_{\pm 0.0}$	$37.5_{\pm 0.0}$	93%
	SAR	$56.8_{\pm 0.2}$	$60.7_{\pm 0.0}$	$61.0_{\pm 0.0}$	93%
	DeYO	$64.3_{\pm 0.3}$	$66.0_{\pm 0.0}$	$66.2_{\pm 0.0}$	89%
	ReCAP	$64.0_{\pm 0.1}$	$65.3_{\pm 0.0}$	$65.6_{\pm 0.0}$	81%
ResNet-50 GN	Source	$31.4_{\pm 0.0}$	$36.4_{\pm 0.0}$	$36.9_{\pm 0.0}$	91%
	SAR	$36.7_{\pm 1.3}$	$46.7_{\pm 0.0}$	$47.1_{\pm 0.0}$	96%
	DeYO	$45.2_{\pm 0.3}$	$51.6_{\pm 0.0}$	$51.9_{\pm 0.0}$	96%
	ReCAP	$46.4_{\pm 0.6}$	$51.8_{\pm 0.0}$	$52.0_{\pm 0.0}$	96%

G Interpolation between within-class and total whitening

Figure 7 reports the accuracy of DeYO combined with T_α on ImageNet-R, whose class frequencies are imbalanced, for $\alpha \in \{0, 0.25, 0.5, 0.75, 1\}$. The accuracy decreases monotonically with α on both backbones, from 62.7 to 59.5 on ViT-B/16 and from 47.8 to 44.2 on ResNet-50 GN. Since no intermediate value of α improves the accuracy over within-class whitening ($\alpha = 0$), we conclude that including the class-mean covariance in the whitening operator degrades the adaptation performance under an imbalanced class distribution such as that of ImageNet-R.

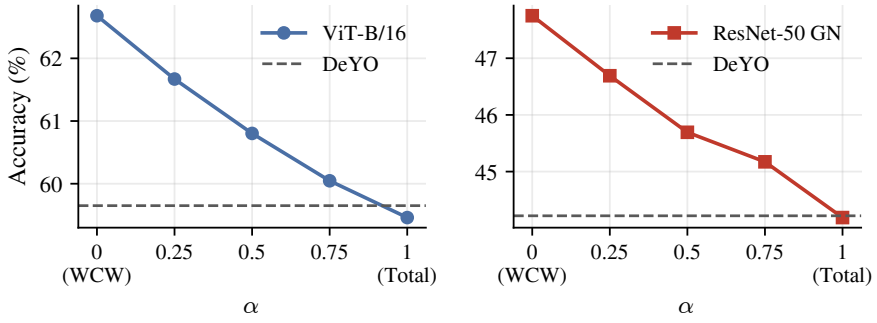


Figure 7: **Accuracy on ImageNet-R against α of T_α.** Here, $\alpha = 0$ corresponds to WCW and $\alpha = 1$ to Total. The dashed line is the accuracy of DeYO without whitening.

H Spectral flatness of all corruptions

Figure 8 repeats the measurement of Figure 2 for every ImageNet-C corruption on ViT-B/16. The median flatness before adaptation lies below the source level for 14 of the 15 corruptions, the median flatness under each of the six existing methods remains below the source level for every corruption, and WCW raises the flatness of nearly every class above its value under any of the existing methods.

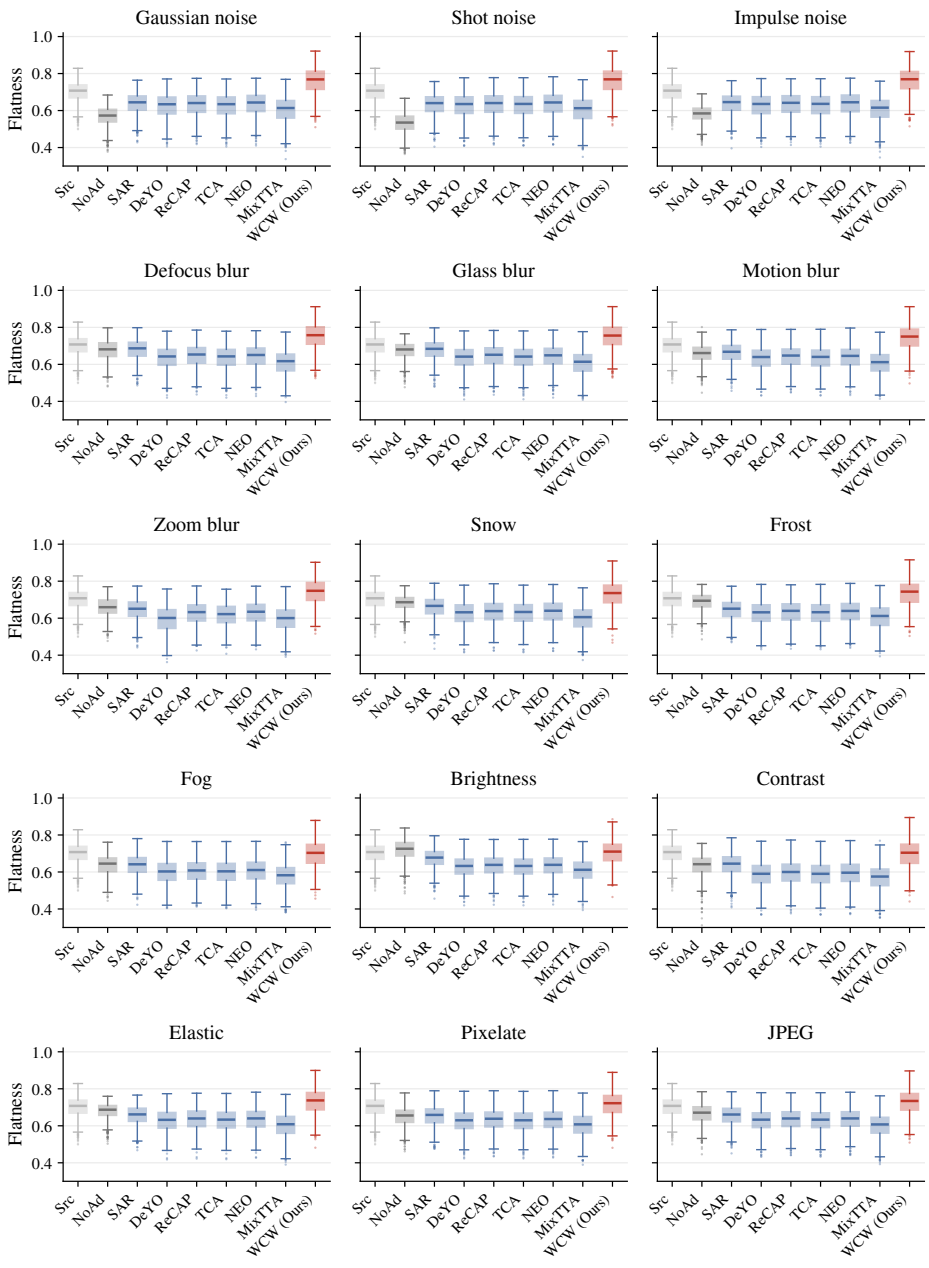


Figure 8: **Spectral flatness of the covariance of each class for the 15 ImageNet-C corruptions (ViT-B/16).** Each panel follows the format of Figure 2; Src and NoAd abbreviate Source and No Adapt.

I PER-CORRUPTION ADAPTATION RESULTS

Tables 8–10 report the per-corruption accuracies for the three ImageNet-C adaptation scenarios that have a corruption axis.

Table 8: ImageNet-C, mild scenario: top-1 accuracy (%) per corruption for ViT-B/16 (top) and ResNet-50 GN (bottom).

Method	Noise			Blur				Weather				Digital				Avg.
	Gauss.	Shot	Impulse	Defocus	Glass	Motion	Zoom	Snow	Frost	Fog	Brit.	Contr.	Elastic	Pixel	JPEG	
ViT-B/16																
Source	16.8$_{\pm0.0}$	12.0$_{\pm0.0}$	16.5$_{\pm0.0}$	29.2$_{\pm0.0}$	23.6$_{\pm0.0}$	33.9$_{\pm0.0}$	27.3$_{\pm0.0}$	15.8$_{\pm0.0}$	26.6$_{\pm0.0}$	47.5$_{\pm0.0}$	55.4$_{\pm0.0}$	44.3$_{\pm0.0}$	31.0$_{\pm0.0}$	45.1$_{\pm0.0}$	49.1$_{\pm0.0}$	31.6$_{\pm0.0}$
+TCA	17.3$_{\pm0.0}$	12.6$_{\pm0.0}$	17.0$_{\pm0.0}$	29.3$_{\pm0.0}$	23.9$_{\pm0.0}$	34.7$_{\pm0.0}$	28.0$_{\pm0.0}$	16.5$_{\pm0.0}$	27.1$_{\pm0.0}$	48.6$_{\pm0.0}$	55.5$_{\pm0.0}$	44.8$_{\pm0.0}$	32.0$_{\pm0.0}$	45.3$_{\pm0.0}$	49.1$_{\pm0.0}$	32.1$_{\pm0.0}$
+NEO	19.7$_{\pm0.0}$	14.7$_{\pm0.0}$	19.2$_{\pm0.0}$	33.3$_{\pm0.0}$	26.7$_{\pm0.0}$	37.1$_{\pm0.0}$	31.1$_{\pm0.0}$	18.8$_{\pm0.0}$	30.1$_{\pm0.0}$	52.7$_{\pm0.0}$	57.4$_{\pm0.0}$	48.2$_{\pm0.0}$	36.2$_{\pm0.0}$	47.5$_{\pm0.0}$	52.0$_{\pm0.0}$	35.0$_{\pm0.0}$
+MixTTA	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
+WCW	**21.3**$_{\pm0.0}$	**16.3**$_{\pm0.0}$	**21.1**$_{\pm0.0}$	**35.3**$_{\pm0.0}$	**28.8**$_{\pm0.0}$	**39.8**$_{\pm0.0}$	**33.9**$_{\pm0.1}$	**20.9**$_{\pm0.0}$	**32.2**$_{\pm0.0}$	**55.4**$_{\pm0.1}$	**58.8**$_{\pm0.1}$	**51.2**$_{\pm0.0}$	**39.1**$_{\pm0.1}$	**49.2**$_{\pm0.0}$	**53.5**$_{\pm0.0}$	**37.1**$_{\pm0.0}$
SAR	46.3$_{\pm0.1}$	45.0$_{\pm0.2}$	46.9$_{\pm0.1}$	52.9$_{\pm0.1}$	50.0$_{\pm0.1}$	55.9$_{\pm0.1}$	51.5$_{\pm0.1}$	57.0$_{\pm0.9}$	53.5$_{\pm2.4}$	66.5$_{\pm0.1}$	74.8$_{\pm0.0}$	64.4$_{\pm0.1}$	55.2$_{\pm0.3}$	66.8$_{\pm0.1}$	64.5$_{\pm0.0}$	56.8$_{\pm0.2}$
+TCA	46.4$_{\pm0.1}$	45.2$_{\pm0.2}$	47.0$_{\pm0.1}$	53.0$_{\pm0.1}$	50.1$_{\pm0.1}$	56.0$_{\pm0.1}$	51.8$_{\pm0.1}$	57.7$_{\pm0.6}$	55.8$_{\pm1.1}$	66.6$_{\pm0.1}$	74.8$_{\pm0.1}$	64.4$_{\pm0.1}$	55.5$_{\pm0.3}$	66.8$_{\pm0.0}$	64.5$_{\pm0.0}$	57.0$_{\pm0.1}$
+NEO	47.0$_{\pm0.2}$	46.0$_{\pm0.1}$	47.8$_{\pm0.1}$	53.8$_{\pm0.1}$	51.6$_{\pm0.0}$	56.7$_{\pm0.1}$	53.2$_{\pm0.1}$	60.0$_{\pm0.1}$	60.3$_{\pm0.2}$	67.7$_{\pm0.1}$	75.3$_{\pm0.0}$	65.1$_{\pm0.1}$	58.4$_{\pm0.0}$	67.3$_{\pm0.0}$	65.5$_{\pm0.0}$	58.4$_{\pm0.0}$
+MixTTA	47.8$_{\pm0.1}$	46.9$_{\pm0.3}$	49.1$_{\pm0.2}$	51.7$_{\pm0.3}$	49.0$_{\pm1.2}$	55.7$_{\pm0.5}$	50.4$_{\pm0.5}$	58.9$_{\pm0.5}$	57.4$_{\pm0.3}$	67.0$_{\pm0.2}$	74.8$_{\pm0.1}$	64.1$_{\pm0.0}$	53.9$_{\pm0.7}$	67.4$_{\pm0.5}$	64.2$_{\pm0.2}$	57.2$_{\pm0.2}$
+WCW	**50.1**$_{\pm0.1}$	**49.3**$_{\pm0.1}$	**50.8**$_{\pm0.0}$	**55.6**$_{\pm0.1}$	**54.4**$_{\pm0.1}$	**59.2**$_{\pm0.1}$	**56.6**$_{\pm0.1}$	**62.3**$_{\pm0.1}$	**61.9**$_{\pm0.1}$	**69.3**$_{\pm0.1}$	**76.1**$_{\pm0.0}$	**66.1**$_{\pm0.1}$	**62.0**$_{\pm0.1}$	**69.2**$_{\pm0.1}$	**67.1**$_{\pm0.1}$	**60.7**$_{\pm0.0}$
DeYO	54.5$_{\pm0.0}$	54.8$_{\pm0.1}$	55.5$_{\pm0.1}$	58.2$_{\pm0.1}$	58.7$_{\pm0.1}$	63.2$_{\pm0.2}$	56.6$_{\pm3.4}$	67.0$_{\pm0.1}$	65.7$_{\pm0.1}$	73.3$_{\pm0.0}$	78.2$_{\pm0.1}$	68.3$_{\pm0.1}$	67.8$_{\pm0.1}$	73.2$_{\pm0.2}$	70.3$_{\pm0.0}$	64.3$_{\pm0.3}$
+TCA	54.5$_{\pm0.0}$	54.9$_{\pm0.1}$	55.5$_{\pm0.1}$	58.3$_{\pm0.1}$	58.7$_{\pm0.1}$	63.2$_{\pm0.1}$	58.0$_{\pm2.1}$	67.1$_{\pm0.0}$	65.7$_{\pm0.1}$	73.3$_{\pm0.0}$	78.2$_{\pm0.1}$	68.3$_{\pm0.1}$	67.8$_{\pm0.1}$	73.2$_{\pm0.2}$	70.3$_{\pm0.0}$	64.5$_{\pm0.2}$
+NEO	54.9$_{\pm0.1}$	55.3$_{\pm0.2}$	55.9$_{\pm0.2}$	58.8$_{\pm0.0}$	59.2$_{\pm0.1}$	63.6$_{\pm0.2}$	61.2$_{\pm0.2}$	67.7$_{\pm0.1}$	66.4$_{\pm0.1}$	73.6$_{\pm0.1}$	78.3$_{\pm0.0}$	68.4$_{\pm0.1}$	68.6$_{\pm0.1}$	73.4$_{\pm0.1}$	70.6$_{\pm0.0}$	65.0$_{\pm0.0}$
+MixTTA	56.0$_{\pm0.2}$	56.6$_{\pm0.1}$	57.0$_{\pm0.0}$	58.2$_{\pm0.3}$	59.3$_{\pm0.3}$	64.6$_{\pm0.3}$	62.2$_{\pm0.1}$	**68.6**$_{\pm0.1}$	**67.1**$_{\pm0.1}$	**74.0**$_{\pm0.2}$	**78.5**$_{\pm0.1}$	68.6$_{\pm0.1}$	68.4$_{\pm0.2}$	**74.2**$_{\pm0.2}$	**71.0**$_{\pm0.0}$	65.6$_{\pm0.1}$
+WCW	**56.3**$_{\pm0.1}$	**57.0**$_{\pm0.0}$	**57.4**$_{\pm0.1}$	**59.6**$_{\pm0.1}$	**60.2**$_{\pm0.1}$	**64.8**$_{\pm0.1}$	**63.4**$_{\pm0.1}$	68.4$_{\pm0.1}$	66.9$_{\pm0.1}$	73.8$_{\pm0.1}$	78.3$_{\pm0.0}$	**68.9**$_{\pm0.1}$	**69.5**$_{\pm0.1}$	73.8$_{\pm0.1}$	**71.0**$_{\pm0.0}$	**66.0**$_{\pm0.0}$
ReCAP	53.9$_{\pm0.0}$	54.2$_{\pm0.1}$	55.0$_{\pm0.0}$	57.9$_{\pm0.0}$	58.2$_{\pm0.1}$	62.6$_{\pm0.1}$	58.8$_{\pm0.6}$	66.4$_{\pm0.0}$	65.2$_{\pm0.1}$	72.9$_{\pm0.0}$	78.0$_{\pm0.1}$	67.9$_{\pm0.1}$	67.0$_{\pm0.1}$	72.7$_{\pm0.1}$	69.9$_{\pm0.1}$	64.0$_{\pm0.1}$
+TCA	53.9$_{\pm0.1}$	54.2$_{\pm0.1}$	55.0$_{\pm0.1}$	58.0$_{\pm0.1}$	58.1$_{\pm0.1}$	62.7$_{\pm0.1}$	59.2$_{\pm0.4}$	66.5$_{\pm0.0}$	65.2$_{\pm0.1}$	72.9$_{\pm0.1}$	77.9$_{\pm0.0}$	67.9$_{\pm0.1}$	67.0$_{\pm0.1}$	72.7$_{\pm0.1}$	69.9$_{\pm0.1}$	64.1$_{\pm0.0}$
+NEO	54.3$_{\pm0.1}$	54.7$_{\pm0.1}$	55.3$_{\pm0.1}$	58.4$_{\pm0.0}$	58.8$_{\pm0.1}$	63.0$_{\pm0.2}$	60.3$_{\pm0.2}$	67.1$_{\pm0.1}$	65.9$_{\pm0.1}$	73.2$_{\pm0.1}$	78.1$_{\pm0.0}$	68.3$_{\pm0.0}$	67.8$_{\pm0.1}$	72.9$_{\pm0.1}$	70.2$_{\pm0.1}$	64.5$_{\pm0.1}$
+MixTTA	55.4$_{\pm0.1}$	55.7$_{\pm0.1}$	56.4$_{\pm0.0}$	57.6$_{\pm0.3}$	58.6$_{\pm0.2}$	63.6$_{\pm0.4}$	60.9$_{\pm0.1}$	**67.9**$_{\pm0.1}$	**66.4**$_{\pm0.2}$	**73.4**$_{\pm0.1}$	**78.2**$_{\pm0.1}$	68.4$_{\pm0.1}$	67.1$_{\pm0.3}$	**73.5**$_{\pm0.1}$	70.4$_{\pm0.1}$	64.9$_{\pm0.1}$
+WCW	**55.7**$_{\pm0.1}$	**56.2**$_{\pm0.0}$	**56.6**$_{\pm0.1}$	**59.1**$_{\pm0.1}$	**59.6**$_{\pm0.1}$	**64.1**$_{\pm0.2}$	**62.5**$_{\pm0.1}$	67.7$_{\pm0.1}$	66.3$_{\pm0.0}$	**73.4**$_{\pm0.1}$	78.1$_{\pm0.1}$	**68.5**$_{\pm0.1}$	**68.6**$_{\pm0.1}$	73.2$_{\pm0.1}$	**70.6**$_{\pm0.0}$	**65.3**$_{\pm0.0}$
ResNet-50 GN																
Source	22.1$_{\pm0.0}$	23.0$_{\pm0.0}$	22.0$_{\pm0.0}$	19.8$_{\pm0.0}$	11.4$_{\pm0.0}$	21.5$_{\pm0.0}$	25.0$_{\pm0.0}$	40.3$_{\pm0.0}$	47.0$_{\pm0.0}$	34.0$_{\pm0.0}$	68.8$_{\pm0.0}$	36.2$_{\pm0.0}$	18.5$_{\pm0.0}$	29.2$_{\pm0.0}$	52.6$_{\pm0.0}$	31.4$_{\pm0.0}$
+TCA	22.9$_{\pm0.0}$	23.6$_{\pm0.0}$	22.7$_{\pm0.0}$	20.3$_{\pm0.0}$	12.1$_{\pm0.0}$	22.8$_{\pm0.0}$	25.8$_{\pm0.0}$	41.6$_{\pm0.0}$	47.5$_{\pm0.0}$	39.0$_{\pm0.0}$	69.1$_{\pm0.0}$	37.6$_{\pm0.0}$	20.4$_{\pm0.0}$	29.7$_{\pm0.0}$	52.9$_{\pm0.0}$	32.5$_{\pm0.0}$
+NEO	24.3$_{\pm0.0}$	25.2$_{\pm0.0}$	24.6$_{\pm0.0}$	22.0$_{\pm0.0}$	13.1$_{\pm0.0}$	24.2$_{\pm0.0}$	27.1$_{\pm0.0}$	42.8$_{\pm0.0}$	48.8$_{\pm0.0}$	48.2$_{\pm0.0}$	69.3$_{\pm0.0}$	38.7$_{\pm0.0}$	21.8$_{\pm0.0}$	31.3$_{\pm0.0}$	53.7$_{\pm0.0}$	34.3$_{\pm0.0}$
+MixTTA	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
+WCW	**26.2**$_{\pm0.0}$	**27.0**$_{\pm0.0}$	**26.2**$_{\pm0.0}$	**23.2**$_{\pm0.0}$	**14.9**$_{\pm0.0}$	**26.6**$_{\pm0.0}$	**29.8**$_{\pm0.0}$	**45.7**$_{\pm0.0}$	**50.3**$_{\pm0.1}$	**49.1**$_{\pm0.0}$	**69.7**$_{\pm0.0}$	**42.1**$_{\pm0.0}$	**26.7**$_{\pm0.1}$	**33.4**$_{\pm0.1}$	**55.4**$_{\pm0.1}$	**36.4**$_{\pm0.0}$
SAR	32.1$_{\pm0.3}$	34.0$_{\pm0.1}$	33.2$_{\pm0.3}$	18.5$_{\pm0.2}$	18.5$_{\pm0.5}$	30.4$_{\pm0.4}$	30.5$_{\pm0.4}$	42.4$_{\pm0.3}$	43.4$_{\pm0.1}$	33.3$_{\pm17.7}$	70.3$_{\pm0.0}$	44.0$_{\pm0.0}$	15.5$_{\pm0.5}$	49.0$_{\pm0.1}$	55.3$_{\pm0.0}$	36.7$_{\pm1.3}$
+TCA	32.6$_{\pm0.3}$	34.5$_{\pm0.1}$	33.7$_{\pm0.3}$	19.3$_{\pm0.2}$	19.2$_{\pm0.6}$	31.2$_{\pm0.4}$	31.5$_{\pm0.4}$	44.4$_{\pm0.2}$	44.4$_{\pm0.1}$	47.5$_{\pm1.1}$	70.5$_{\pm0.0}$	44.6$_{\pm0.0}$	18.3$_{\pm0.3}$	49.6$_{\pm0.1}$	55.6$_{\pm0.0}$	38.5$_{\pm0.1}$
+NEO	35.4$_{\pm0.2}$	37.2$_{\pm0.1}$	36.7$_{\pm0.1}$	23.9$_{\pm0.1}$	24.1$_{\pm0.0}$	33.6$_{\pm0.1}$	35.1$_{\pm0.2}$	48.3$_{\pm0.1}$	47.1$_{\pm0.1}$	53.5$_{\pm0.1}$	70.8$_{\pm0.1}$	46.6$_{\pm0.1}$	35.6$_{\pm0.1}$	51.8$_{\pm0.1}$	56.7$_{\pm0.1}$	42.4$_{\pm0.0}$
+MixTTA	33.7$_{\pm0.4}$	36.0$_{\pm0.5}$	34.9$_{\pm0.2}$	18.8$_{\pm0.1}$	20.3$_{\pm0.2}$	31.5$_{\pm0.2}$	31.9$_{\pm0.3}$	41.8$_{\pm0.5}$	43.6$_{\pm0.5}$	34.6$_{\pm17.1}$	70.5$_{\pm0.1}$	44.4$_{\pm0.2}$	13.9$_{\pm0.4}$	50.5$_{\pm0.1}$	55.5$_{\pm0.0}$	37.5$_{\pm1.3}$
+WCW	**39.0**$_{\pm0.1}$	**40.5**$_{\pm0.2}$	**39.8**$_{\pm0.0}$	**32.1**$_{\pm0.1}$	**31.1**$_{\pm0.1}$	**38.4**$_{\pm0.1}$	**41.5**$_{\pm0.0}$	**51.4**$_{\pm0.1}$	**51.5**$_{\pm0.2}$	**55.2**$_{\pm0.1}$	**71.8**$_{\pm0.0}$	**50.5**$_{\pm0.1}$	**43.3**$_{\pm0.2}$	**55.8**$_{\pm0.0}$	**58.6**$_{\pm0.0}$	**46.7**$_{\pm0.0}$
DeYO	40.8$_{\pm0.2}$	42.9$_{\pm0.1}$	41.9$_{\pm0.1}$	21.9$_{\pm0.1}$	23.9$_{\pm0.1}$	38.3$_{\pm0.0}$	34.2$_{\pm4.1}$	50.6$_{\pm0.3}$	49.1$_{\pm0.1}$	55.8$_{\pm0.2}$	72.9$_{\pm0.1}$	50.0$_{\pm0.1}$	42.0$_{\pm0.3}$	56.0$_{\pm0.2}$	57.9$_{\pm0.0}$	45.2$_{\pm0.3}$
+TCA	41.0$_{\pm0.3}$	43.0$_{\pm0.2}$	42.1$_{\pm0.1}$	22.9$_{\pm0.1}$	24.6$_{\pm0.1}$	38.7$_{\pm0.0}$	35.9$_{\pm3.3}$	51.4$_{\pm0.2}$	49.7$_{\pm0.1}$	57.1$_{\pm0.1}$	72.9$_{\pm0.1}$	50.4$_{\pm0.1}$	42.8$_{\pm0.3}$	56.6$_{\pm0.1}$	58.2$_{\pm0.1}$	45.8$_{\pm0.2}$
+NEO	43.6$_{\pm0.1}$	45.5$_{\pm0.1}$	44.7$_{\pm0.0}$	33.2$_{\pm0.3}$	31.7$_{\pm0.2}$	42.1$_{\pm0.2}$	43.5$_{\pm0.1}$	54.6$_{\pm0.1}$	53.1$_{\pm0.1}$	**60.8**$_{\pm0.1}$	**73.5**$_{\pm0.0}$	52.9$_{\pm0.0}$	48.9$_{\pm0.1}$	59.3$_{\pm0.1}$	59.6$_{\pm0.1}$	49.8$_{\pm0.1}$
+MixTTA	42.9$_{\pm0.4}$	44.9$_{\pm0.3}$	44.1$_{\pm0.5}$	23.7$_{\pm0.3}$	25.2$_{\pm0.1}$	40.4$_{\pm0.2}$	36.3$_{\pm1.3}$	52.4$_{\pm0.1}$	50.5$_{\pm0.4}$	20.2$_{\pm26.9}$	73.2$_{\pm0.1}$	51.1$_{\pm0.2}$	46.0$_{\pm0.4}$	57.8$_{\pm0.2}$	58.0$_{\pm0.1}$	44.4$_{\pm1.8}$
+WCW	**44.4**$_{\pm0.1}$	**46.1**$_{\pm0.2}$	**45.5**$_{\pm0.2}$	**36.1**$_{\pm0.2}$	**35.8**$_{\pm0.3}$	**44.8**$_{\pm0.1}$	**48.5**$_{\pm0.1}$	**55.3**$_{\pm0.1}$	**54.2**$_{\pm0.1}$	60.4$_{\pm0.2}$	73.0$_{\pm0.0}$	**54.9**$_{\pm0.1}$	**53.2**$_{\pm0.1}$	**61.0**$_{\pm0.0}$	**61.1**$_{\pm0.1}$	**51.6**$_{\pm0.0}$
ReCAP	41.8$_{\pm0.1}$	43.6$_{\pm0.1}$	43.0$_{\pm0.1}$	24.4$_{\pm0.6}$	25.4$_{\pm0.4}$	40.2$_{\pm0.3}$	40.7$_{\pm0.2}$	51.9$_{\pm0.2}$	50.4$_{\pm0.0}$	57.0$_{\pm0.3}$	73.1$_{\pm0.0}$	51.0$_{\pm0.2}$	37.9$_{\pm9.7}$	57.2$_{\pm0.2}$	58.4$_{\pm0.1}$	46.4$_{\pm0.6}$
+TCA	42.1$_{\pm0.1}$	43.8$_{\pm0.1}$	43.2$_{\pm0.2}$	25.5$_{\pm0.4}$	26.2$_{\pm0.3}$	40.6$_{\pm0.3}$	41.7$_{\pm0.2}$	52.5$_{\pm0.1}$	50.7$_{\pm0.1}$	58.1$_{\pm0.2}$	73.1$_{\pm0.0}$	51.3$_{\pm0.1}$	39.1$_{\pm8.8}$	57.6$_{\pm0.1}$	58.7$_{\pm0.1}$	47.0$_{\pm0.6}$
+NEO	44.1$_{\pm0.0}$	45.9$_{\pm0.1}$	45.2$_{\pm0.0}$	34.2$_{\pm0.3}$	32.6$_{\pm0.3}$	43.1$_{\pm0.2}$	44.6$_{\pm0.0}$	55.1$_{\pm0.1}$	53.7$_{\pm0.1}$	**61.2**$_{\pm0.2}$	**73.6**$_{\pm0.1}$	53.4$_{\pm0.1}$	50.1$_{\pm0.2}$	60.0$_{\pm0.0}$	59.9$_{\pm0.1}$	50.4$_{\pm0.1}$
+MixTTA	43.5$_{\pm0.4}$	45.4$_{\pm0.2}$	44.4$_{\pm0.4}$	25.2$_{\pm0.3}$	26.7$_{\pm0.2}$	41.9$_{\pm0.3}$	40.9$_{\pm1.7}$	53.0$_{\pm0.2}$	51.3$_{\pm0.1}$	59.2$_{\pm1.1}$	73.4$_{\pm0.1}$	51.8$_{\pm0.1}$	46.9$_{\pm0.4}$	58.6$_{\pm0.2}$	58.5$_{\pm0.2}$	48.0$_{\pm0.1}$
+WCW	**44.5**$_{\pm0.1}$	**46.3**$_{\pm0.1}$	**45.6**$_{\pm0.3}$	**36.2**$_{\pm0.1}$	**36.0**$_{\pm0.2}$	**45.1**$_{\pm0.2}$	**48.8**$_{\pm0.1}$	**55.4**$_{\pm0.1}$	**54.1**$_{\pm0.0}$	60.5$_{\pm0.4}$	73.0$_{\pm0.0}$	**55.1**$_{\pm0.1}$	**53.3**$_{\pm0.2}$	**61.3**$_{\pm0.1}$	**61.2**$_{\pm0.0}$	**51.8**$_{\pm0.0}$

Table 9: ImageNet-C, label shifts: top-1 accuracy (%) per corruption for ViT-B/16 (top) and ResNet-50 GN (bottom).

Method	Noise			Blur				Weather				Digital				Avg.
	Gauss.	Shot	Impulse	Defocus	Glass	Motion	Zoom	Snow	Frost	Fog	Brit.	Contr.	Elastic	Pixel	JPEG	
ViT-B/16																
Source	$16.7_{\pm 0.0}$	$11.9_{\pm 0.0}$	$16.4_{\pm 0.1}$	$29.3_{\pm 0.0}$	$23.6_{\pm 0.0}$	$34.0_{\pm 0.0}$	$27.3_{\pm 0.0}$	$15.8_{\pm 0.1}$	$26.5_{\pm 0.0}$	$47.5_{\pm 0.2}$	$55.2_{\pm 0.2}$	$44.2_{\pm 0.0}$	$30.9_{\pm 0.0}$	$45.1_{\pm 0.0}$	$48.9_{\pm 0.1}$	$31.6_{\pm 0.0}$
+TCA	$17.5_{\pm 0.0}$	$12.7_{\pm 0.0}$	$17.1_{\pm 0.1}$	$29.7_{\pm 0.0}$	$23.9_{\pm 0.0}$	$35.3_{\pm 0.0}$	$28.3_{\pm 0.0}$	$16.7_{\pm 0.0}$	$27.4_{\pm 0.0}$	$49.0_{\pm 0.2}$	$55.5_{\pm 0.2}$	$45.1_{\pm 0.1}$	$32.3_{\pm 0.0}$	$45.6_{\pm 0.0}$	$49.3_{\pm 0.0}$	$32.4_{\pm 0.0}$
+NEO	$19.5_{\pm 0.0}$	$14.5_{\pm 0.1}$	$19.0_{\pm 0.1}$	$33.2_{\pm 0.0}$	$26.6_{\pm 0.0}$	$37.1_{\pm 0.1}$	$30.9_{\pm 0.0}$	$18.7_{\pm 0.0}$	$30.0_{\pm 0.0}$	$52.4_{\pm 0.1}$	$57.0_{\pm 0.1}$	$48.0_{\pm 0.1}$	$35.8_{\pm 0.0}$	$47.4_{\pm 0.0}$	$51.8_{\pm 0.0}$	$34.8_{\pm 0.0}$
+MixTTA	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
+WCW	$\mathbf{21.4}_{\pm 0.1}$	$\mathbf{16.6}_{\pm 0.1}$	$\mathbf{21.2}_{\pm 0.1}$	$\mathbf{35.8}_{\pm 0.0}$	$\mathbf{29.0}_{\pm 0.1}$	$\mathbf{40.2}_{\pm 0.1}$	$\mathbf{34.3}_{\pm 0.1}$	$\mathbf{21.1}_{\pm 0.1}$	$\mathbf{32.5}_{\pm 0.1}$	$\mathbf{55.8}_{\pm 0.1}$	$\mathbf{58.8}_{\pm 0.2}$	$\mathbf{51.5}_{\pm 0.1}$	$\mathbf{39.4}_{\pm 0.2}$	$\mathbf{49.5}_{\pm 0.1}$	$\mathbf{53.6}_{\pm 0.1}$	$\mathbf{37.4}_{\pm 0.1}$
SAR	$50.0_{\pm 0.1}$	$49.1_{\pm 0.2}$	$50.6_{\pm 0.0}$	$55.3_{\pm 0.2}$	$54.2_{\pm 0.3}$	$59.1_{\pm 0.2}$	$54.6_{\pm 0.1}$	$57.2_{\pm 6.1}$	$45.6_{\pm 2.8}$	$69.7_{\pm 0.3}$	$76.3_{\pm 0.0}$	$66.3_{\pm 0.1}$	$60.8_{\pm 0.4}$	$69.8_{\pm 0.2}$	$66.9_{\pm 0.1}$	$59.0_{\pm 0.5}$
+TCA	$50.2_{\pm 0.1}$	$49.3_{\pm 0.2}$	$50.8_{\pm 0.1}$	$55.6_{\pm 0.2}$	$54.4_{\pm 0.3}$	$59.4_{\pm 0.2}$	$55.2_{\pm 0.1}$	$59.7_{\pm 3.3}$	$53.9_{\pm 0.1}$	$69.9_{\pm 0.3}$	$76.3_{\pm 0.0}$	$66.3_{\pm 0.2}$	$61.3_{\pm 0.4}$	$69.8_{\pm 0.1}$	$67.0_{\pm 0.2}$	$59.9_{\pm 0.3}$
+NEO	$50.4_{\pm 0.1}$	$49.8_{\pm 0.1}$	$51.3_{\pm 0.1}$	$56.0_{\pm 0.2}$	$55.1_{\pm 0.3}$	$59.6_{\pm 0.1}$	$56.1_{\pm 0.2}$	$63.3_{\pm 0.2}$	$62.8_{\pm 0.1}$	$70.3_{\pm 0.2}$	$76.4_{\pm 0.1}$	$66.3_{\pm 0.4}$	$63.1_{\pm 0.1}$	$69.9_{\pm 0.1}$	$67.4_{\pm 0.2}$	$61.2_{\pm 0.1}$
+MixTTA	$52.0_{\pm 0.1}$	$51.2_{\pm 0.1}$	$52.7_{\pm 0.2}$	$54.6_{\pm 0.1}$	$54.6_{\pm 0.7}$	$59.5_{\pm 0.3}$	$54.8_{\pm 0.2}$	$63.1_{\pm 0.3}$	$58.4_{\pm 2.2}$	$70.4_{\pm 0.3}$	$76.3_{\pm 0.1}$	$66.3_{\pm 0.2}$	$60.1_{\pm 0.8}$	$70.5_{\pm 0.2}$	$67.5_{\pm 0.1}$	$60.8_{\pm 0.2}$
+WCW	$\mathbf{53.4}_{\pm 0.1}$	$\mathbf{53.1}_{\pm 0.1}$	$\mathbf{54.3}_{\pm 0.1}$	$\mathbf{57.5}_{\pm 0.1}$	$\mathbf{57.7}_{\pm 0.1}$	$\mathbf{61.9}_{\pm 0.0}$	$\mathbf{59.7}_{\pm 0.2}$	$\mathbf{65.4}_{\pm 0.1}$	$\mathbf{64.6}_{\pm 0.1}$	$\mathbf{71.8}_{\pm 0.2}$	$\mathbf{77.2}_{\pm 0.0}$	$\mathbf{67.6}_{\pm 0.1}$	$\mathbf{66.4}_{\pm 0.2}$	$\mathbf{71.6}_{\pm 0.2}$	$\mathbf{69.1}_{\pm 0.0}$	$\mathbf{63.4}_{\pm 0.0}$
DeYO	$54.5_{\pm 0.1}$	$55.1_{\pm 0.0}$	$55.8_{\pm 0.3}$	$58.0_{\pm 0.3}$	$58.8_{\pm 0.0}$	$63.8_{\pm 0.1}$	$61.3_{\pm 0.2}$	$67.9_{\pm 0.3}$	$66.1_{\pm 0.1}$	$73.0_{\pm 0.2}$	$78.0_{\pm 0.1}$	$66.8_{\pm 0.1}$	$69.1_{\pm 0.2}$	$73.6_{\pm 0.1}$	$70.6_{\pm 0.1}$	$64.8_{\pm 0.0}$
+TCA	$54.6_{\pm 0.1}$	$55.3_{\pm 0.1}$	$55.9_{\pm 0.3}$	$58.0_{\pm 0.4}$	$59.0_{\pm 0.2}$	$63.9_{\pm 0.2}$	$61.7_{\pm 0.4}$	$68.1_{\pm 0.2}$	$66.3_{\pm 0.1}$	$73.1_{\pm 0.1}$	$78.1_{\pm 0.1}$	$66.8_{\pm 0.2}$	$69.2_{\pm 0.3}$	$73.7_{\pm 0.1}$	$70.6_{\pm 0.1}$	$65.0_{\pm 0.1}$
+NEO	$54.7_{\pm 0.1}$	$55.5_{\pm 0.1}$	$55.9_{\pm 0.4}$	$58.0_{\pm 0.2}$	$59.1_{\pm 0.2}$	$63.9_{\pm 0.0}$	$62.3_{\pm 0.1}$	$68.2_{\pm 0.2}$	$66.3_{\pm 0.1}$	$72.9_{\pm 0.2}$	$77.9_{\pm 0.1}$	$66.7_{\pm 0.4}$	$69.4_{\pm 0.2}$	$73.6_{\pm 0.1}$	$70.5_{\pm 0.1}$	$65.0_{\pm 0.0}$
+MixTTA	$56.0_{\pm 0.5}$	$56.3_{\pm 0.3}$	$56.9_{\pm 0.2}$	$57.9_{\pm 0.1}$	$59.3_{\pm 0.1}$	$64.7_{\pm 0.2}$	$62.8_{\pm 0.2}$	$68.9_{\pm 0.2}$	$67.2_{\pm 0.3}$	$\mathbf{73.9}_{\pm 0.0}$	$\mathbf{78.4}_{\pm 0.1}$	$67.4_{\pm 0.3}$	$69.2_{\pm 0.3}$	$\mathbf{74.2}_{\pm 0.1}$	$71.1_{\pm 0.1}$	$65.6_{\pm 0.1}$
+WCW	$\mathbf{56.6}_{\pm 0.1}$	$\mathbf{57.4}_{\pm 0.1}$	$\mathbf{57.7}_{\pm 0.1}$	$\mathbf{59.4}_{\pm 0.2}$	$\mathbf{60.4}_{\pm 0.1}$	$\mathbf{65.7}_{\pm 0.1}$	$\mathbf{64.6}_{\pm 0.1}$	$\mathbf{69.4}_{\pm 0.1}$	$\mathbf{67.4}_{\pm 0.2}$	$73.8_{\pm 0.2}$	$78.1_{\pm 0.0}$	$\mathbf{67.6}_{\pm 0.1}$	$\mathbf{70.4}_{\pm 0.1}$	$\mathbf{74.2}_{\pm 0.0}$	$\mathbf{71.3}_{\pm 0.2}$	$\mathbf{66.3}_{\pm 0.0}$
ReCAP	$54.4_{\pm 0.3}$	$54.9_{\pm 0.2}$	$55.6_{\pm 0.1}$	$58.1_{\pm 0.2}$	$58.9_{\pm 0.1}$	$63.6_{\pm 0.1}$	$61.0_{\pm 0.4}$	$67.8_{\pm 0.1}$	$66.1_{\pm 0.2}$	$73.2_{\pm 0.2}$	$77.9_{\pm 0.1}$	$66.9_{\pm 0.0}$	$68.7_{\pm 0.0}$	$73.4_{\pm 0.1}$	$70.4_{\pm 0.2}$	$64.7_{\pm 0.1}$
+TCA	$54.6_{\pm 0.2}$	$55.0_{\pm 0.2}$	$55.7_{\pm 0.1}$	$58.2_{\pm 0.2}$	$59.0_{\pm 0.1}$	$63.7_{\pm 0.0}$	$46.5_{\pm 21.5}$	$67.9_{\pm 0.1}$	$66.1_{\pm 0.2}$	$73.2_{\pm 0.1}$	$78.0_{\pm 0.1}$	$67.2_{\pm 0.2}$	$68.8_{\pm 0.0}$	$73.5_{\pm 0.2}$	$70.4_{\pm 0.1}$	$63.9_{\pm 1.5}$
+NEO	$54.7_{\pm 0.2}$	$55.4_{\pm 0.0}$	$55.8_{\pm 0.2}$	$58.2_{\pm 0.2}$	$59.1_{\pm 0.1}$	$63.8_{\pm 0.2}$	$61.5_{\pm 0.4}$	$67.9_{\pm 0.0}$	$66.2_{\pm 0.0}$	$73.2_{\pm 0.1}$	$77.9_{\pm 0.1}$	$66.8_{\pm 0.2}$	$69.0_{\pm 0.1}$	$73.4_{\pm 0.0}$	$70.4_{\pm 0.1}$	$64.9_{\pm 0.1}$
+MixTTA	$55.8_{\pm 0.4}$	$56.1_{\pm 0.2}$	$56.8_{\pm 0.1}$	$57.8_{\pm 0.0}$	$59.3_{\pm 0.2}$	$64.9_{\pm 0.0}$	$63.5_{\pm 0.5}$	$\mathbf{69.0}_{\pm 0.2}$	$67.2_{\pm 0.4}$	$\mathbf{73.8}_{\pm 0.2}$	$\mathbf{78.4}_{\pm 0.1}$	$\mathbf{67.9}_{\pm 0.1}$	$69.3_{\pm 0.3}$	$\mathbf{74.3}_{\pm 0.1}$	$71.0_{\pm 0.1}$	$65.7_{\pm 0.1}$
+WCW	$\mathbf{56.3}_{\pm 0.2}$	$\mathbf{56.8}_{\pm 0.1}$	$\mathbf{57.4}_{\pm 0.1}$	$\mathbf{59.3}_{\pm 0.3}$	$\mathbf{60.2}_{\pm 0.1}$	$\mathbf{65.2}_{\pm 0.2}$	$\mathbf{63.8}_{\pm 0.2}$	$68.9_{\pm 0.0}$	$\mathbf{67.3}_{\pm 0.1}$	$\mathbf{73.8}_{\pm 0.1}$	$78.1_{\pm 0.2}$	$\mathbf{67.9}_{\pm 0.1}$	$\mathbf{69.8}_{\pm 0.1}$	$73.9_{\pm 0.1}$	$\mathbf{71.1}_{\pm 0.1}$	$\mathbf{66.0}_{\pm 0.1}$
ResNet-50 GN																
Source	$22.1_{\pm 0.0}$	$23.0_{\pm 0.1}$	$22.1_{\pm 0.1}$	$19.9_{\pm 0.0}$	$11.4_{\pm 0.0}$	$21.4_{\pm 0.1}$	$25.0_{\pm 0.1}$	$40.3_{\pm 0.1}$	$47.0_{\pm 0.1}$	$34.0_{\pm 0.1}$	$68.8_{\pm 0.1}$	$36.3_{\pm 0.0}$	$18.5_{\pm 0.0}$	$29.2_{\pm 0.0}$	$52.6_{\pm 0.0}$	$31.4_{\pm 0.0}$
+TCA	$23.2_{\pm 0.0}$	$24.0_{\pm 0.1}$	$23.0_{\pm 0.1}$	$20.5_{\pm 0.1}$	$12.3_{\pm 0.0}$	$23.0_{\pm 0.1}$	$26.1_{\pm 0.1}$	$42.4_{\pm 0.1}$	$47.6_{\pm 0.1}$	$39.2_{\pm 0.0}$	$69.2_{\pm 0.1}$	$38.2_{\pm 0.0}$	$20.9_{\pm 0.1}$	$29.7_{\pm 0.0}$	$53.4_{\pm 0.0}$	$32.8_{\pm 0.0}$
+NEO	$24.2_{\pm 0.0}$	$25.1_{\pm 0.1}$	$24.5_{\pm 0.1}$	$21.9_{\pm 0.1}$	$13.0_{\pm 0.1}$	$24.0_{\pm 0.0}$	$27.0_{\pm 0.1}$	$42.7_{\pm 0.1}$	$48.7_{\pm 0.1}$	$48.0_{\pm 0.1}$	$69.1_{\pm 0.1}$	$38.6_{\pm 0.0}$	$21.8_{\pm 0.0}$	$31.1_{\pm 0.0}$	$53.6_{\pm 0.0}$	$34.2_{\pm 0.0}$
+MixTTA	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
+WCW	$\mathbf{26.5}_{\pm 0.0}$	$\mathbf{27.4}_{\pm 0.1}$	$\mathbf{26.6}_{\pm 0.1}$	$\mathbf{23.7}_{\pm 0.1}$	$\mathbf{15.3}_{\pm 0.1}$	$\mathbf{27.2}_{\pm 0.1}$	$\mathbf{30.1}_{\pm 0.1}$	$\mathbf{46.2}_{\pm 0.2}$	$\mathbf{50.5}_{\pm 0.2}$	$\mathbf{50.1}_{\pm 0.1}$	$\mathbf{69.8}_{\pm 0.1}$	$\mathbf{42.6}_{\pm 0.1}$	$\mathbf{27.4}_{\pm 0.1}$	$\mathbf{33.9}_{\pm 0.1}$	$\mathbf{55.7}_{\pm 0.1}$	$\mathbf{36.9}_{\pm 0.0}$
SAR	$36.2_{\pm 0.3}$	$38.4_{\pm 0.2}$	$37.7_{\pm 0.2}$	$18.4_{\pm 0.4}$	$20.6_{\pm 1.2}$	$33.6_{\pm 0.3}$	$30.4_{\pm 3.6}$	$25.0_{\pm 1.8}$	$44.9_{\pm 0.6}$	$20.9_{\pm 20.9}$	$71.4_{\pm 0.1}$	$47.0_{\pm 0.0}$	$6.7_{\pm 1.0}$	$52.4_{\pm 0.3}$	$56.5_{\pm 0.0}$	$36.0_{\pm 1.2}$
+TCA	$36.8_{\pm 0.3}$	$39.1_{\pm 0.4}$	$38.4_{\pm 0.2}$	$19.8_{\pm 0.4}$	$21.6_{\pm 1.2}$	$34.4_{\pm 0.3}$	$32.0_{\pm 4.0}$	$31.6_{\pm 2.3}$	$46.7_{\pm 0.3}$	$39.3_{\pm 16.9}$	$71.7_{\pm 0.1}$	$47.8_{\pm 0.1}$	$8.6_{\pm 0.9}$	$53.2_{\pm 0.2}$	$57.0_{\pm 0.1}$	$38.5_{\pm 1.2}$
+NEO	$38.9_{\pm 0.3}$	$40.8_{\pm 0.2}$	$40.0_{\pm 0.0}$	$25.7_{\pm 0.3}$	$26.7_{\pm 0.1}$	$36.6_{\pm 0.4}$	$38.9_{\pm 0.1}$	$50.4_{\pm 0.2}$	$48.9_{\pm 0.5}$	$55.9_{\pm 0.1}$	$71.7_{\pm 0.1}$	$49.0_{\pm 0.0}$	$40.5_{\pm 0.1}$	$54.6_{\pm 0.1}$	$57.5_{\pm 0.1}$	$45.1_{\pm 0.1}$
+MixTTA	$37.9_{\pm 0.3}$	$40.1_{\pm 0.2}$	$38.7_{\pm 0.3}$	$18.5_{\pm 0.7}$	$20.8_{\pm 1.7}$	$34.5_{\pm 0.2}$	$35.4_{\pm 0.7}$	$21.5_{\pm 2.6}$	$45.3_{\pm 0.5}$	$35.3_{\pm 23.6}$	$71.7_{\pm 0.1}$	$47.4_{\pm 0.1}$	$5.8_{\pm 0.3}$	$53.5_{\pm 0.2}$	$56.5_{\pm 0.0}$	$37.5_{\pm 1.6}$
+WCW	$\mathbf{42.4}_{\pm 0.2}$	$\mathbf{44.3}_{\pm 0.1}$	$\mathbf{43.5}_{\pm 0.4}$	$\mathbf{34.6}_{\pm 0.1}$	$\mathbf{33.4}_{\pm 0.1}$	$\mathbf{42.2}_{\pm 0.2}$	$\mathbf{45.4}_{\pm 0.1}$	$\mathbf{53.9}_{\pm 0.1}$	$\mathbf{53.3}_{\pm 0.2}$	$\mathbf{58.4}_{\pm 0.0}$	$\mathbf{72.6}_{\pm 0.1}$	$\mathbf{53.2}_{\pm 0.2}$	$\mathbf{50.2}_{\pm 0.4}$	$\mathbf{58.3}_{\pm 0.2}$	$\mathbf{60.0}_{\pm 0.1}$	$\mathbf{49.7}_{\pm 0.1}$
DeYO	$42.6_{\pm 0.1}$	$45.0_{\pm 0.2}$	$41.7_{\pm 2.7}$	$21.9_{\pm 0.8}$	$8.8_{\pm 11.0}$	$41.1_{\pm 0.4}$	$14.1_{\pm 6.9}$	$52.5_{\pm 0.2}$	$50.0_{\pm 0.3}$	$20.3_{\pm 27.5}$	$73.2_{\pm 0.2}$	$52.5_{\pm 0.1}$	$36.6_{\pm 13.8}$	$59.5_{\pm 0.3}$	$59.3_{\pm 0.1}$	$41.3_{\pm 3.0}$
+TCA	$42.9_{\pm 0.2}$	$45.3_{\pm 0.3}$	$44.0_{\pm 0.2}$	$23.3_{\pm 1.0}$	$9.5_{\pm 11.4}$	$41.7_{\pm 0.3}$	$17.2_{\pm 7.5}$	$53.1_{\pm 0.1}$	$49.3_{\pm 2.9}$	$24.7_{\pm 24.8}$	$73.3_{\pm 0.2}$	$52.8_{\pm 0.1}$	$40.2_{\pm 10.1}$	$60.2_{\pm 0.2}$	$59.7_{\pm 0.2}$	$42.5_{\pm 2.3}$
+NEO	$44.4_{\pm 0.0}$	$46.5_{\pm 0.3}$	$45.3_{\pm 0.1}$	$33.9_{\pm 0.2}$	$33.3_{\pm 0.3}$	$44.1_{\pm 0.1}$	$44.8_{\pm 0.4}$	$56.0_{\pm 0.1}$	$54.2_{\pm 0.2}$	$\mathbf{61.8}_{\pm 0.2}$	$\mathbf{73.5}_{\pm 0.2}$	$54.6_{\pm 0.1}$	$52.0_{\pm 0.4}$	$61.8_{\pm 0.2}$	$60.7_{\pm 0.2}$	$51.1_{\pm 0.1}$
+MixTTA	$43.6_{\pm 0.5}$	$45.9_{\pm 0.3}$	$44.8_{\pm 0.1}$	$18.5_{\pm 5.7}$	$8.7_{\pm 10.8}$	$42.9_{\pm 0.5}$	$7.7_{\pm 2.3}$	$53.3_{\pm 0.6}$	$51.0_{\pm 0.2}$	$20.4_{\pm 28.0}$	$73.3_{\pm 0.2}$	$52.9_{\pm 0.1}$	$49.2_{\pm 0.4}$	$60.5_{\pm 0.4}$	$59.1_{\pm 0.2}$	$42.1_{\pm 2.7}$
+WCW	$\mathbf{45.0}_{\pm 0.4}$	$\mathbf{46.7}_{\pm 0.3}$	$\mathbf{45.7}_{\pm 0.3}$	$\mathbf{36.5}_{\pm 0.4}$	$\mathbf{37.0}_{\pm 0.2}$	$\mathbf{46.6}_{\pm 0.3}$	$\mathbf{50.5}_{\pm 0.2}$	$\mathbf{57.1}_{\pm 0.1}$	$\mathbf{55.0}_{\pm 0.2}$	$61.4_{\pm 0.5}$	$72.8_{\pm 0.1}$	$\mathbf{56.0}_{\pm 0.3}$	$\mathbf{55.6}_{\pm 0.2}$	$\mathbf{63.1}_{\pm 0.1}$	$\mathbf{61.6}_{\pm 0.0}$	$\mathbf{52.7}_{\pm 0.0}$
ReCAP	$43.4_{\pm 0.3}$	$45.5_{\pm 0.2}$	$43.7_{\pm 0.6}$	$24.3_{\pm 0.3}$	$9.1_{\pm 11.4}$	$42.8_{\pm 0.4}$	$32.7_{\pm 5.1}$	$53.6_{\pm 0.4}$	$52.0_{\pm 0.1}$	$20.4_{\pm 27.9}$	$73.4_{\pm 0.1}$	$53.1_{\pm 0.0}$	$24.9_{\pm 19.2}$	$60.6_{\pm 0.2}$	$59.9_{\pm 0.3}$	$42.6_{\pm 3.9}$
+TCA	$43.5_{\pm 0.2}$	$45.6_{\pm 0.0}$	$44.1_{\pm 0.1}$	$25.4_{\pm 0.5}$	$10.1_{\pm 12.3}$	$43.2_{\pm 0.4}$	$35.4_{\pm 4.2}$	$54.0_{\pm 0.2}$	$52.4_{\pm 0.2}$	$24.3_{\pm 25.6}$	$73.4_{\pm 0.2}$	$53.7_{\pm 0.0}$	$28.5_{\pm 19.5}$	$61.0_{\pm 0.2}$	$60.3_{\pm 0.2}$	$43.7_{\pm 3.8}$
+NEO	$\mathbf{44.7}_{\pm 0.1}$	$\mathbf{46.5}_{\pm 0.2}$	$\mathbf{45.4}_{\pm 0.1}$	$34.1_{\pm 0.1}$	$33.4_{\pm 0.4}$	$44.6_{\pm 0.4}$	$46.5_{\pm 0.3}$	$56.5_{\pm 0.2}$	$54.3_{\pm 0.2}$	$\mathbf{62.5}_{\pm 0.3}$	$\mathbf{73.6}_{\pm 0.1}$	$55.0_{\pm 0.2}$	$52.8_{\pm 0.5}$	$62.3_{\pm 0.3}$	$61.0_{\pm 0.2}$	$51.5_{\pm 0.1}$
+MixTTA	$43.7_{\pm 0.4}$	$46.2_{\pm 0.2}$	$44.5_{\pm 0.2}$	$24.4_{\pm 1.0}$	$16.8_{\pm 11.5}$	$43.7_{\pm 0.8}$	$18.8_{\pm 13.0}$	$54.2_{\pm 0.7}$	$51.9_{\pm 0.3}$	$20.7_{\pm 28.3}$	$73.4_{\pm 0.2}$	$53.6_{\pm 0.1}$	$39.9_{\pm 15.1}$	$61.7_{\pm 0.5}$	$59.9_{\pm 0.4}$	$43.5_{\pm 2.3}$
+WCW	$44.4_{\pm 0.3}$	$46.1_{\pm 0.2}$	$\mathbf{45.4}_{\pm 0.1}$	$\mathbf{36.2}_{\pm 0.1}$	$\mathbf{36.7}_{\pm 0.4}$	$\mathbf{46.5}_{\pm 0.3}$	$\mathbf{50.5}_{\pm 0.1}$	$\mathbf{56.9}_{\pm 0.1}$	$\mathbf{55.0}_{\pm 0.4}$	$60.8_{\pm 0.8}$	$72.7_{\pm 0.1}$	$\mathbf{55.8}_{\pm 0.2}$	$\mathbf{55.3}_{\pm 0.1}$	$\mathbf{63.1}_{\pm 0.2}$	$\mathbf{61.7}_{\pm 0.1}$	$\mathbf{52.5}_{\pm 0.1}$

Table 10: ImageNet-C, batch size 1: top-1 accuracy (%) per corruption for ViT-B/16 (top) and ResNet-50 GN (bottom).

Method	Noise			Blur				Weather				Digital				Avg.
	Gauss.	Shot	Impulse	Defocus	Glass	Motion	Zoom	Snow	Frost	Fog	Brit.	Contr.	Elastic	Pixel	JPEG	
ViT-B/16																
Source	16.8$_{\pm0.0}$	12.0$_{\pm0.0}$	16.5$_{\pm0.0}$	29.2$_{\pm0.0}$	23.6$_{\pm0.0}$	33.9$_{\pm0.0}$	27.3$_{\pm0.0}$	15.8$_{\pm0.0}$	26.6$_{\pm0.0}$	47.5$_{\pm0.0}$	55.4$_{\pm0.0}$	44.3$_{\pm0.0}$	31.0$_{\pm0.0}$	45.1$_{\pm0.0}$	49.1$_{\pm0.0}$	31.6$_{\pm0.0}$
+TCA	17.3$_{\pm0.0}$	12.6$_{\pm0.0}$	17.0$_{\pm0.0}$	29.3$_{\pm0.0}$	23.9$_{\pm0.0}$	34.7$_{\pm0.0}$	28.0$_{\pm0.0}$	16.5$_{\pm0.0}$	27.1$_{\pm0.0}$	48.6$_{\pm0.0}$	55.5$_{\pm0.0}$	44.8$_{\pm0.0}$	32.0$_{\pm0.0}$	45.3$_{\pm0.0}$	49.1$_{\pm0.0}$	32.1$_{\pm0.0}$
+NEO	19.7$_{\pm0.0}$	14.7$_{\pm0.0}$	19.2$_{\pm0.0}$	33.3$_{\pm0.0}$	26.7$_{\pm0.0}$	37.1$_{\pm0.0}$	31.1$_{\pm0.0}$	18.8$_{\pm0.0}$	30.1$_{\pm0.0}$	52.7$_{\pm0.0}$	57.4$_{\pm0.0}$	48.2$_{\pm0.0}$	36.2$_{\pm0.0}$	47.5$_{\pm0.0}$	52.0$_{\pm0.0}$	35.0$_{\pm0.0}$
+MixTTA	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
+WCW	**21.3**$_{\pm0.0}$	**16.3**$_{\pm0.0}$	**21.1**$_{\pm0.0}$	**35.3**$_{\pm0.0}$	**28.8**$_{\pm0.1}$	**39.7**$_{\pm0.0}$	**33.9**$_{\pm0.1}$	**20.9**$_{\pm0.0}$	**32.2**$_{\pm0.0}$	**55.4**$_{\pm0.1}$	**58.7**$_{\pm0.1}$	**51.2**$_{\pm0.0}$	**39.1**$_{\pm0.1}$	**49.2**$_{\pm0.0}$	**53.5**$_{\pm0.0}$	**37.1**$_{\pm0.0}$
SAR	45.6$_{\pm0.1}$	43.2$_{\pm0.1}$	46.3$_{\pm0.1}$	53.6$_{\pm0.0}$	50.5$_{\pm0.1}$	57.6$_{\pm0.1}$	53.1$_{\pm0.2}$	58.7$_{\pm0.1}$	54.8$_{\pm2.9}$	68.9$_{\pm0.2}$	75.4$_{\pm0.7}$	65.6$_{\pm0.1}$	58.1$_{\pm0.1}$	69.0$_{\pm0.1}$	66.3$_{\pm0.1}$	57.8$_{\pm0.2}$
+TCA	45.7$_{\pm0.1}$	43.3$_{\pm0.2}$	46.4$_{\pm0.1}$	53.6$_{\pm0.1}$	50.6$_{\pm0.2}$	57.7$_{\pm0.0}$	53.4$_{\pm0.2}$	59.0$_{\pm0.1}$	56.5$_{\pm2.2}$	69.0$_{\pm0.1}$	75.4$_{\pm0.6}$	65.6$_{\pm0.1}$	58.3$_{\pm0.1}$	69.0$_{\pm0.0}$	66.4$_{\pm0.1}$	58.0$_{\pm0.2}$
+NEO	48.2$_{\pm0.1}$	46.7$_{\pm0.1}$	49.0$_{\pm0.1}$	55.1$_{\pm0.0}$	53.0$_{\pm0.1}$	58.8$_{\pm0.1}$	55.2$_{\pm0.2}$	61.5$_{\pm0.1}$	61.7$_{\pm0.2}$	69.5$_{\pm0.8}$	75.2$_{\pm0.1}$	66.4$_{\pm0.1}$	61.2$_{\pm0.2}$	69.7$_{\pm0.0}$	67.3$_{\pm0.1}$	59.9$_{\pm0.1}$
+MixTTA	47.1$_{\pm0.3}$	44.9$_{\pm0.3}$	48.4$_{\pm0.3}$	52.0$_{\pm0.4}$	49.5$_{\pm1.2}$	57.0$_{\pm0.4}$	51.6$_{\pm0.4}$	59.5$_{\pm0.7}$	58.7$_{\pm0.5}$	69.0$_{\pm0.1}$	**76.0**$_{\pm0.0}$	65.4$_{\pm0.3}$	56.0$_{\pm0.6}$	67.4$_{\pm1.5}$	66.1$_{\pm0.1}$	57.9$_{\pm0.2}$
+WCW	**49.4**$_{\pm0.1}$	**47.7**$_{\pm0.3}$	**50.3**$_{\pm0.1}$	**55.9**$_{\pm0.1}$	**54.4**$_{\pm0.2}$	**60.3**$_{\pm0.1}$	**57.7**$_{\pm0.1}$	**62.4**$_{\pm0.1}$	**62.7**$_{\pm0.0}$	**71.0**$_{\pm0.1}$	75.8$_{\pm0.4}$	**67.0**$_{\pm0.1}$	**63.6**$_{\pm0.1}$	**70.7**$_{\pm0.0}$	**68.3**$_{\pm0.0}$	**61.1**$_{\pm0.1}$
DeYO	55.2$_{\pm0.1}$	55.6$_{\pm0.1}$	56.3$_{\pm0.0}$	58.9$_{\pm0.0}$	59.4$_{\pm0.2}$	64.4$_{\pm0.1}$	46.5$_{\pm3.0}$	68.1$_{\pm0.1}$	66.5$_{\pm0.1}$	73.8$_{\pm0.0}$	78.4$_{\pm0.1}$	68.3$_{\pm0.0}$	69.0$_{\pm0.1}$	73.9$_{\pm0.1}$	70.9$_{\pm0.0}$	64.3$_{\pm0.2}$
+TCA	55.2$_{\pm0.1}$	55.6$_{\pm0.2}$	56.3$_{\pm0.0}$	58.9$_{\pm0.0}$	59.4$_{\pm0.1}$	64.3$_{\pm0.1}$	49.5$_{\pm5.6}$	68.1$_{\pm0.1}$	66.5$_{\pm0.0}$	73.8$_{\pm0.1}$	78.4$_{\pm0.1}$	68.3$_{\pm0.1}$	69.0$_{\pm0.1}$	73.9$_{\pm0.1}$	70.9$_{\pm0.1}$	64.5$_{\pm0.4}$
+NEO	55.7$_{\pm0.1}$	56.3$_{\pm0.2}$	56.8$_{\pm0.1}$	59.2$_{\pm0.1}$	59.9$_{\pm0.1}$	64.6$_{\pm0.1}$	62.2$_{\pm0.4}$	68.6$_{\pm0.1}$	67.0$_{\pm0.1}$	74.0$_{\pm0.1}$	78.4$_{\pm0.1}$	68.5$_{\pm0.0}$	69.5$_{\pm0.0}$	74.1$_{\pm0.1}$	71.1$_{\pm0.0}$	65.7$_{\pm0.0}$
+MixTTA	56.2$_{\pm0.0}$	56.8$_{\pm0.2}$	57.2$_{\pm0.1}$	58.6$_{\pm0.2}$	59.9$_{\pm0.3}$	**65.6**$_{\pm0.3}$	63.9$_{\pm0.4}$	**69.3**$_{\pm0.1}$	**67.6**$_{\pm0.1}$	**74.5**$_{\pm0.1}$	**78.8**$_{\pm0.1}$	**69.0**$_{\pm0.1}$	69.4$_{\pm0.2}$	**74.7**$_{\pm0.1}$	**71.5**$_{\pm0.1}$	66.2$_{\pm0.1}$
+WCW	**56.7**$_{\pm0.1}$	**57.3**$_{\pm0.1}$	**57.8**$_{\pm0.0}$	**60.0**$_{\pm0.1}$	**60.7**$_{\pm0.0}$	**65.6**$_{\pm0.1}$	**64.4**$_{\pm0.1}$	69.1$_{\pm0.2}$	67.4$_{\pm0.0}$	74.2$_{\pm0.1}$	78.5$_{\pm0.1}$	68.9$_{\pm0.1}$	**70.1**$_{\pm0.1}$	74.2$_{\pm0.1}$	71.4$_{\pm0.0}$	**66.4**$_{\pm0.0}$
ReCAP	56.0$_{\pm0.1}$	56.8$_{\pm0.2}$	57.2$_{\pm0.2}$	59.1$_{\pm0.3}$	60.3$_{\pm0.0}$	65.5$_{\pm0.0}$	47.3$_{\pm14.3}$	69.2$_{\pm0.1}$	67.3$_{\pm0.1}$	74.0$_{\pm0.1}$	78.5$_{\pm0.0}$	65.7$_{\pm2.9}$	70.1$_{\pm0.1}$	74.4$_{\pm0.1}$	71.3$_{\pm0.0}$	64.8$_{\pm1.0}$
+TCA	56.0$_{\pm0.2}$	56.9$_{\pm0.1}$	57.2$_{\pm0.1}$	59.2$_{\pm0.1}$	60.3$_{\pm0.1}$	65.4$_{\pm0.0}$	57.2$_{\pm5.8}$	69.1$_{\pm0.1}$	67.4$_{\pm0.1}$	74.0$_{\pm0.1}$	78.5$_{\pm0.1}$	66.2$_{\pm2.0}$	70.1$_{\pm0.1}$	74.4$_{\pm0.1}$	71.4$_{\pm0.1}$	65.6$_{\pm0.4}$
+NEO	56.4$_{\pm0.1}$	57.3$_{\pm0.1}$	57.7$_{\pm0.2}$	59.5$_{\pm0.1}$	60.5$_{\pm0.1}$	65.6$_{\pm0.0}$	60.8$_{\pm4.4}$	69.6$_{\pm0.2}$	67.6$_{\pm0.1}$	74.2$_{\pm0.0}$	78.6$_{\pm0.0}$	67.8$_{\pm0.3}$	70.5$_{\pm0.1}$	74.6$_{\pm0.1}$	71.5$_{\pm0.1}$	66.1$_{\pm0.3}$
+MixTTA	56.6$_{\pm0.1}$	57.5$_{\pm0.3}$	57.9$_{\pm0.1}$	58.6$_{\pm0.3}$	60.6$_{\pm0.1}$	**66.3**$_{\pm0.1}$	**66.5**$_{\pm0.1}$	**70.1**$_{\pm0.1}$	**68.3**$_{\pm0.1}$	**74.5**$_{\pm0.1}$	**78.8**$_{\pm0.1}$	67.9$_{\pm0.1}$	**70.8**$_{\pm0.0}$	**75.0**$_{\pm0.0}$	**71.8**$_{\pm0.1}$	66.8$_{\pm0.1}$
+WCW	**57.3**$_{\pm0.1}$	**58.0**$_{\pm0.1}$	**58.3**$_{\pm0.1}$	**60.3**$_{\pm0.1}$	**61.1**$_{\pm0.0}$	**66.3**$_{\pm0.0}$	65.5$_{\pm0.1}$	69.6$_{\pm0.0}$	67.9$_{\pm0.1}$	74.4$_{\pm0.1}$	78.5$_{\pm0.0}$	**68.7**$_{\pm0.1}$	70.6$_{\pm0.1}$	74.6$_{\pm0.0}$	71.6$_{\pm0.0}$	**66.9**$_{\pm0.0}$
ResNet-50 GN																
Source	22.1$_{\pm0.0}$	23.0$_{\pm0.0}$	22.0$_{\pm0.0}$	19.8$_{\pm0.0}$	11.4$_{\pm0.0}$	21.5$_{\pm0.0}$	25.0$_{\pm0.0}$	40.3$_{\pm0.0}$	47.0$_{\pm0.0}$	34.0$_{\pm0.0}$	68.8$_{\pm0.0}$	36.2$_{\pm0.0}$	18.5$_{\pm0.0}$	29.2$_{\pm0.0}$	52.6$_{\pm0.0}$	31.4$_{\pm0.0}$
+TCA	22.9$_{\pm0.0}$	23.6$_{\pm0.0}$	22.7$_{\pm0.0}$	20.3$_{\pm0.0}$	12.1$_{\pm0.0}$	22.8$_{\pm0.0}$	25.8$_{\pm0.0}$	41.6$_{\pm0.0}$	47.5$_{\pm0.0}$	39.0$_{\pm0.0}$	69.1$_{\pm0.0}$	37.6$_{\pm0.0}$	20.4$_{\pm0.0}$	29.7$_{\pm0.0}$	52.9$_{\pm0.0}$	32.5$_{\pm0.0}$
+NEO	24.3$_{\pm0.0}$	25.2$_{\pm0.0}$	24.6$_{\pm0.0}$	22.1$_{\pm0.0}$	13.1$_{\pm0.0}$	24.2$_{\pm0.0}$	27.1$_{\pm0.0}$	42.8$_{\pm0.0}$	48.8$_{\pm0.0}$	48.2$_{\pm0.0}$	69.3$_{\pm0.0}$	38.7$_{\pm0.0}$	21.8$_{\pm0.0}$	31.3$_{\pm0.0}$	53.7$_{\pm0.0}$	34.3$_{\pm0.0}$
+MixTTA	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
+WCW	**26.2**$_{\pm0.0}$	**27.0**$_{\pm0.0}$	**26.2**$_{\pm0.0}$	**23.2**$_{\pm0.0}$	**14.9**$_{\pm0.0}$	**26.6**$_{\pm0.1}$	**29.8**$_{\pm0.0}$	**45.8**$_{\pm0.1}$	**50.3**$_{\pm0.1}$	**49.1**$_{\pm0.0}$	**69.7**$_{\pm0.0}$	**42.1**$_{\pm0.0}$	**26.7**$_{\pm0.1}$	**33.5**$_{\pm0.1}$	**55.4**$_{\pm0.1}$	**36.4**$_{\pm0.0}$
SAR	28.7$_{\pm0.1}$	31.0$_{\pm0.1}$	29.2$_{\pm0.2}$	18.4$_{\pm0.2}$	15.2$_{\pm0.0}$	28.7$_{\pm0.2}$	30.5$_{\pm0.1}$	44.9$_{\pm0.4}$	44.4$_{\pm0.2}$	33.8$_{\pm0.4}$	72.0$_{\pm0.1}$	44.7$_{\pm0.0}$	14.4$_{\pm0.4}$	47.7$_{\pm0.4}$	56.3$_{\pm0.1}$	36.0$_{\pm0.0}$
+TCA	29.3$_{\pm0.2}$	31.3$_{\pm0.0}$	29.8$_{\pm0.2}$	19.1$_{\pm0.2}$	16.0$_{\pm0.1}$	29.6$_{\pm0.2}$	31.4$_{\pm0.1}$	46.4$_{\pm0.3}$	45.3$_{\pm0.2}$	41.2$_{\pm0.5}$	72.0$_{\pm0.0}$	45.3$_{\pm0.0}$	17.1$_{\pm0.3}$	48.1$_{\pm0.4}$	56.6$_{\pm0.1}$	37.2$_{\pm0.0}$
+NEO	**37.6**$_{\pm0.2}$	**39.9**$_{\pm0.2}$	**39.0**$_{\pm0.2}$	25.1$_{\pm0.2}$	23.9$_{\pm0.3}$	36.1$_{\pm0.1}$	39.4$_{\pm0.1}$	51.5$_{\pm0.1}$	49.9$_{\pm0.1}$	**56.8**$_{\pm0.0}$	**72.7**$_{\pm0.1}$	49.4$_{\pm0.1}$	41.0$_{\pm0.2}$	**54.8**$_{\pm0.1}$	58.2$_{\pm0.0}$	45.0$_{\pm0.0}$
+MixTTA	29.9$_{\pm0.1}$	32.6$_{\pm0.5}$	30.6$_{\pm0.2}$	18.5$_{\pm0.0}$	17.0$_{\pm0.1}$	29.9$_{\pm0.2}$	31.5$_{\pm0.2}$	45.4$_{\pm0.3}$	44.6$_{\pm0.2}$	33.9$_{\pm2.6}$	72.2$_{\pm0.1}$	44.8$_{\pm0.3}$	13.5$_{\pm0.6}$	49.3$_{\pm0.2}$	56.3$_{\pm0.0}$	36.7$_{\pm0.3}$
+WCW	37.1$_{\pm0.0}$	38.9$_{\pm0.1}$	37.7$_{\pm0.2}$	**28.5**$_{\pm0.1}$	**24.0**$_{\pm0.2}$	**37.6**$_{\pm0.0}$	**40.4**$_{\pm0.1}$	**51.9**$_{\pm0.1}$	**52.2**$_{\pm0.1}$	56.7$_{\pm0.1}$	72.6$_{\pm0.0}$	**50.8**$_{\pm0.1}$	**42.6**$_{\pm0.2}$	52.5$_{\pm0.2}$	**59.1**$_{\pm0.0}$	**45.5**$_{\pm0.0}$
DeYO	42.7$_{\pm0.2}$	45.1$_{\pm0.2}$	44.1$_{\pm0.1}$	23.0$_{\pm0.2}$	24.1$_{\pm0.1}$	41.7$_{\pm0.1}$	26.2$_{\pm10.9}$	54.0$_{\pm0.0}$	52.0$_{\pm0.1}$	39.6$_{\pm26.3}$	73.3$_{\pm0.0}$	53.4$_{\pm0.1}$	47.7$_{\pm0.4}$	60.1$_{\pm0.3}$	59.8$_{\pm0.1}$	45.8$_{\pm1.7}$
+TCA	42.8$_{\pm0.1}$	45.2$_{\pm0.2}$	44.1$_{\pm0.2}$	23.9$_{\pm0.2}$	24.8$_{\pm0.2}$	42.0$_{\pm0.2}$	28.7$_{\pm10.2}$	54.4$_{\pm0.1}$	52.3$_{\pm0.0}$	27.9$_{\pm21.9}$	73.3$_{\pm0.1}$	53.7$_{\pm0.1}$	48.2$_{\pm0.4}$	60.3$_{\pm0.1}$	60.0$_{\pm0.1}$	45.5$_{\pm0.8}$
+NEO	**45.5**$_{\pm0.1}$	**47.6**$_{\pm0.1}$	**46.6**$_{\pm0.0}$	35.5$_{\pm0.3}$	35.1$_{\pm0.1}$	**45.7**$_{\pm0.4}$	44.9$_{\pm1.0}$	**57.7**$_{\pm0.2}$	**55.0**$_{\pm0.1}$	**63.2**$_{\pm0.1}$	**73.9**$_{\pm0.1}$	**56.0**$_{\pm0.0}$	54.2$_{\pm0.1}$	**63.2**$_{\pm0.1}$	**61.5**$_{\pm0.1}$	**52.4**$_{\pm0.1}$
+MixTTA	44.1$_{\pm0.6}$	46.5$_{\pm0.2}$	45.2$_{\pm0.4}$	22.9$_{\pm0.2}$	25.2$_{\pm0.2}$	43.7$_{\pm0.2}$	30.2$_{\pm1.7}$	55.1$_{\pm0.4}$	52.5$_{\pm0.2}$	21.0$_{\pm27.7}$	73.4$_{\pm0.1}$	54.0$_{\pm0.2}$	51.2$_{\pm0.4}$	61.6$_{\pm0.2}$	59.9$_{\pm0.1}$	45.8$_{\pm1.8}$
+WCW	44.4$_{\pm0.2}$	46.5$_{\pm0.1}$	45.6$_{\pm0.2}$	**35.9**$_{\pm0.2}$	**35.4**$_{\pm0.4}$	45.6$_{\pm0.1}$	**49.2**$_{\pm0.1}$	56.7$_{\pm0.1}$	54.6$_{\pm0.1}$	60.1$_{\pm0.4}$	73.1$_{\pm0.0}$	55.8$_{\pm0.1}$	**54.3**$_{\pm0.0}$	62.3$_{\pm0.0}$	**61.5**$_{\pm0.1}$	52.1$_{\pm0.1}$
ReCAP	43.4$_{\pm0.1}$	45.8$_{\pm0.1}$	44.4$_{\pm0.1}$	26.8$_{\pm1.2}$	26.7$_{\pm0.9}$	43.4$_{\pm0.1}$	41.8$_{\pm3.9}$	55.8$_{\pm0.3}$	52.9$_{\pm0.1}$	40.8$_{\pm27.0}$	73.5$_{\pm0.0}$	54.7$_{\pm0.1}$	40.6$_{\pm13.8}$	61.6$_{\pm0.1}$	60.7$_{\pm0.1}$	47.5$_{\pm2.5}$
+TCA	43.5$_{\pm0.3}$	45.9$_{\pm0.1}$	44.7$_{\pm0.1}$	27.7$_{\pm1.2}$	27.3$_{\pm0.9}$	43.7$_{\pm0.1}$	43.8$_{\pm1.0}$	56.3$_{\pm0.3}$	53.2$_{\pm0.1}$	60.1$_{\pm0.3}$	73.5$_{\pm0.1}$	54.8$_{\pm0.1}$	43.1$_{\pm11.3}$	61.7$_{\pm0.1}$	60.9$_{\pm0.1}$	49.3$_{\pm0.6}$
+NEO	**45.5**$_{\pm0.3}$	**47.7**$_{\pm0.1}$	**46.4**$_{\pm0.2}$	**36.0**$_{\pm0.1}$	**36.2**$_{\pm0.3}$	**47.4**$_{\pm0.2}$	48.1$_{\pm0.3}$	**58.8**$_{\pm0.2}$	**55.7**$_{\pm0.2}$	**64.2**$_{\pm0.0}$	**74.0**$_{\pm0.1}$	**57.0**$_{\pm0.1}$	**56.0**$_{\pm0.2}$	**64.2**$_{\pm0.1}$	**62.1**$_{\pm0.1}$	**53.3**$_{\pm0.0}$
+MixTTA	44.5$_{\pm0.3}$	46.6$_{\pm0.1}$	45.4$_{\pm0.3}$	26.4$_{\pm0.3}$	28.5$_{\pm0.5}$	45.9$_{\pm0.2}$	32.9$_{\pm10.0}$	57.0$_{\pm0.3}$	53.7$_{\pm0.2}$	21.7$_{\pm28.5}$	73.6$_{\pm0.1}$	55.3$_{\pm0.2}$	53.9$_{\pm0.4}$	63.1$_{\pm0.1}$	60.9$_{\pm0.0}$	47.3$_{\pm2.2}$
+WCW	44.3$_{\pm0.1}$	46.5$_{\pm0.3}$	45.3$_{\pm0.4}$	**36.0**$_{\pm0.1}$	35.5$_{\pm0.5}$	46.1$_{\pm0.3}$	**49.8**$_{\pm0.1}$	57.3$_{\pm0.2}$	54.6$_{\pm0.2}$	60.6$_{\pm0.1}$	73.0$_{\pm0.0}$	56.0$_{\pm0.1}$	54.8$_{\pm0.3}$	62.6$_{\pm0.1}$	61.6$_{\pm0.1}$	52.3$_{\pm0.0}$