# Supplementary Materials for CAFA: Class-Aware Feature Alignment for Test-Time Adaptation

This supplementary material presents details about experimental settings, additional experimental results, and visualizations of our method, which are not included in the main paper due to the page limit. Section 1 elaborates on the details about baselines, and Section 2 describes additional experimental results including ablation studies. Finally, Section 3 presents the t-SNE visualizations of CAFA and TENT [10].

## 1. Baselines

**Baselines** For fair comparisons, we consider the following baselines in our experiments:

- Source: Evaluating the pretrained network on the test data without any adaptation.
- Test-time normalization (BN) [7] updates the batch normalization statistics [3] on test data at test time.
- Pseudo label (PL) [5] utilizes the predicted label as a label for optimizing the main task loss during testing. To be specific, we optimize the model by minimizing the cross-entropy loss with the pseudo labels.
- Test-time entropy minimization (TENT) [10] updates the modulation parameters (*i.e.,* $\beta$ and $\gamma$) of batch normalization [3] layers in the network by minimizing entropy on test data.
- Efficient anti-forgetting test-time adaptation (EATA) [8] applies efficient sample selection to filter out redundant samples for adaptation and regularizes important weights using the Fisher matrix to prevent catastrophic forgetting.
- Marginal entropy minimization with one test point (MEMO) [11] adapts a model to a single test point using test-time augmentation and minimizing marginal entropy.
- Feature Restoration (FR-Online) [2] pre-calculates the approximated training feature distributions and adapts a model to test domains by aligning the target feature distributions to the pre-obtained training feature distributions.
- Test-time template adjuster (T3A) [4] calculates a pseudo-prototype vector for each class on test time and classifies images using the distance between each instance and pseudo-prototype vectors.
- Contrastive test-time adaptation (AdaContrast) [1] applies contrastive learning along with online refinement of pseudo labels during test time.
- Test-time training (TTT) [9] first pre-trains the model with a self-supervision loss (*i.e.,* rotation prediction) and adapts the pretrained model to test domains by minimizing the self-supervision loss as a proxy of the main task loss function.
- Test-time feature alignment (TFA-Online) [6] aligns the source and target distributions by matching the first- and second-order statistics of outputs from both the penultimate layer and self-supervised task branch.
- Test-time training++ (TTT++-Online) [6] updates the model by jointly aligning the first- and second-order statistics between the source and target distributions and optimizing the proxy loss (*i.e.,* contrastive loss for self-supervision task) at test time.

## 2. Additional Experimental Results

| Methods | Inference Time (ms) | FPS |
|---|---|---|
| Source | 17.82 | 77.97 |
| BN | 17.45 | 57.31 |
| TENT | 40.47 | 24.71 |
| CAFA | 41.65 | 24.01 |

Table 1. Measured inference time (ms) and FPS for each method. We measure the inference time using a NVIDIA A100 GPU with an image resolution of 224×224. The number is an averaged value of 300 trials.

### 2.1. Inference Time

To show the efficiency of CAFA, we measure the inference time on a single NVIDIA A100 GPU and average the inference time over 300 trials with an image resolution of $224 \times 224$. Our model shows around 24.01 frames per second (FPS). Since CAFA and TENT optimize model parameters by minimizing the loss using gradient descent, those methods have a longer inference time than Source and BN methods. We believe that such a result demonstrates that CAFA not only improves the test time adaptation performance but also maintains a reasonable level of efficiency.

## 2.2. Further Ablation Studies

| Effectiveness of Intra-/Inter-Class Dist. | | Updating Batch Norm. vs Full Parameters | | Tied vs Class-wise Covariance | |
|---|---|---|---|---|---|
| Source | 29.14 | Source | 29.14 | Source | 29.14 |
| Global FA | 19.12 | CAFA-Full | 12.66 | CAFA-Tied | 12.47 |
| Intra-Class Dist. | 13.02 | **CAFA** | **12.13** | **CAFA** | **12.13** |
| **CAFA** | **12.13** | | | | |
| Source | 60.35 | Source | 60.35 | Source | 60.35 |
| Global FA | 51.41 | CAFA-Full | 38.31 | CAFA-Tied | 38.19 |
| Intra-Class Dist. | 41.51 | **CAFA** | **37.31** | **CAFA** | **37.31** |
| **CAFA** | **37.31** | | | | |

Table 2. Our ablation results on the CIFAR10-C (upper group) and CIFAR100-C (lower group) datasets. The left group shows the effectiveness of considering intra- and inter-class distances, the middle group presents the comparison of updating the batch normalization parameters and full parameters in the feature extractor, and the right group describes the effects of using a tied covariance.

**Ablation studies on the CIFAR100-C dataset** Along with the ablation studies in the main paper, we apply the same variants of our methods to the CIFAR100-C dataset. Table 2 shows the ablation studies on both CIFAR10-C (upper group) and CIFAR100-C (lower group) datasets. Overall, we observe the similar results to the ablation studies in our main paper.

| Methods | Classification Error (%) | | |
|---|---|---|---|
| | CIFAR10-C | CIFAR100-C | ImageNet-C |
| Source | 29.14 | 60.35 | 81.99 |
| CAFA-Variance | 12.46 | 37.46 | 55.47 |
| **CAFA** | **12.13** | **37.31** | **55.24** |

Table 3. Effectiveness of using variance in the CIFAR10-C, CIFAR100-C, and ImageNet-C datasets.

**Effectiveness of using variance** Another variant of our loss is using variances of each class-conditional Gaussian distribution instead of using the full covariance. While using variances does not consider the covariance between feature dimensions, it is still distinct for each class-conditional Gaussian distribution. As reported in the upper group of Table 3, adopting class-wise variances also improve the source model significantly. That is, variances can represent the class-conditional Gaussian distributions reasonably well. However, using full covariance reaches the top performance in all three datasets.

## 2.3. Different Corruption Severity Levels

We compare CAFA with the baselines on the CIFAR10-C and CIFAR100-C datasets with different severity levels. As reported in Tables 4-7, CAFA outperforms baselines on all severity levels of the CIFAR10-C and CIFAR100-C datasets. Furthermore, we present the classification errors along with independent trials using different random seeds in Table 8. As shown, CAFA shows minimal performance variation considering the small standard deviation.

# 3. Visualizations

For the qualitative analysis, we visualize the representation space of test samples from our method and TENT [10] by using the t-SNE algorithm. Fig. 1 shows the t-SNE results on different corruption types in the CIFAR10-C dataset, and Fig. 2 illustrates the change of representation space of test samples from our method as adaptation proceeds. As shown, the representations of test samples are well-separated, and they desirably converge in a class-wise manner as adaptation proceeds.

| Method | Gaus. | Shot | Impu. | Defo. | Glas. | Moti. | Zoom | Snow | Fros. | Fog | Brig. | Cont. | Elas. | Pixe. | Jpeg. | Average |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Source | 43.31 | 34.34 | 43.78 | 8.32 | 52.34 | 16.72 | 8.12 | 19.31 | 20.07 | 13.02 | **5.88** | 7.45 | 13.04 | 26.45 | 17.12 | 21.95 |
| BN | 15.97 | 13.14 | 22.03 | 8.30 | 25.85 | 12.42 | 7.23 | 16.99 | 11.70 | 13.07 | 7.49 | 7.49 | 13.57 | 9.25 | 12.38 | 13.13 |
| PL | 15.81 | 13.03 | 21.66 | 8.21 | 25.62 | 12.23 | 7.18 | 16.84 | 11.64 | 12.88 | 7.55 | 7.39 | 13.59 | 9.27 | 12.28 | 13.01 |
| FR-Online† | 15.93 | 13.15 | 21.87 | 8.27 | 25.61 | 12.38 | 7.26 | 16.98 | 11.72 | 13.05 | 7.52 | 7.50 | 13.56 | 9.26 | 12.37 | 13.10 |
| TFA-Online† | 14.69 | 12.29 | 18.03 | 7.57 | 22.84 | 11.19 | 6.59 | 14.77 | 10.48 | 10.62 | 6.68 | 6.89 | 11.90 | 9.08 | 11.03 | 11.64 |
| TTT++-Online† | 15.20 | 12.57 | 17.33 | 7.61 | 23.07 | 10.72 | 6.62 | 13.31 | 10.63 | 9.87 | 6.14 | **6.29** | 12.13 | 8.95 | 11.92 | 11.49 |
| EATA† | 15.52 | 12.73 | 21.08 | 8.16 | 24.72 | 12.07 | 7.12 | 16.50 | 11.54 | 12.62 | 7.42 | 7.41 | 13.34 | 9.07 | 12.04 | 12.76 |
| TENT† | 14.51 | 11.98 | 19.12 | 7.69 | 22.81 | 10.96 | 7.10 | 15.26 | 11.18 | 11.03 | 7.05 | 7.03 | 12.69 | 8.75 | 11.64 | 11.92 |
| **CAFA (Ours)** | **12.73** | **10.51** | **16.71** | **6.66** | **20.34** | **9.73** | **6.07** | **12.82** | **9.51** | **8.98** | 6.14 | 6.35 | **11.25** | **7.76** | **10.31** | **10.39** |
| Source | 76.96 | 69.67 | 79.91 | 32.50 | 82.34 | 46.32 | 32.75 | 50.96 | 51.77 | 44.18 | **26.44** | 32.44 | 41.08 | 52.03 | 46.34 | 51.05 |
| BN | 44.93 | 40.95 | 52.39 | 31.68 | 57.40 | 37.53 | 30.28 | 45.78 | 38.16 | 41.63 | 30.30 | 31.13 | 40.37 | 32.93 | 37.22 | 39.51 |
| PL | 44.54 | 40.42 | 51.50 | 31.63 | 56.51 | 37.05 | 29.95 | 45.38 | 37.78 | 40.93 | 30.16 | 30.82 | 40.04 | 32.56 | 36.85 | 39.07 |
| FR-Online† | 44.92 | 40.99 | 52.39 | 31.72 | 57.38 | 37.55 | 30.33 | 42.01 | 38.14 | 41.62 | 30.32 | 31.17 | 40.36 | 32.90 | 37.19 | 39.27 |
| TFA-Online† | 42.39 | 37.82 | 48.26 | 30.03 | 54.24 | 34.70 | 28.41 | 41.59 | 35.60 | 36.74 | 28.27 | 29.78 | 36.50 | 31.93 | 35.62 | 36.79 |
| TTT++-Online† | 41.16 | 37.15 | 47.38 | 28.74 | 53.18 | 32.81 | 27.95 | 40.65 | 34.85 | 34.30 | 27.34 | 27.83 | 35.41 | 30.59 | 34.99 | 35.62 |
| EATA† | 42.74 | 38.76 | 48.48 | 30.56 | 53.87 | 35.45 | 29.31 | 43.17 | 36.66 | 38.38 | 29.30 | 30.04 | 38.25 | 31.75 | 35.72 | 37.50 |
| TENT† | 41.28 | 37.10 | 46.18 | 29.06 | 51.82 | 33.62 | 28.08 | 41.18 | 35.33 | 35.01 | 27.68 | 28.48 | 36.12 | 30.63 | 34.40 | 35.73 |
| **CAFA (Ours)** | **39.11** | **35.71** | **44.17** | **27.83** | **49.22** | **32.16** | **27.56** | **39.36** | **34.05** | **33.22** | 26.61 | **27.49** | **34.97** | **29.68** | **33.47** | **34.31** |

Table 4. Classification error (%) on the CIFAR10-C (upper group) and CIFAR100-C (lower group) datasets with severity level 4 corruptions. † denotes the results obtained from the official codes.

| Method | Gaus. | Shot | Impu. | Defo. | Glas. | Moti. | Zoom | Snow | Fros. | Fog | Brig. | Cont. | Elas. | Pixe. | Jpeg. | Average |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Source | 36.96 | 28.00 | 26.86 | **5.73** | 35.53 | 16.68 | 7.54 | 16.89 | 18.54 | 8.98 | **5.64** | 6.47 | **7.69** | 13.10 | 15.54 | 16.68 |
| BN | 13.75 | 11.97 | 16.01 | 7.39 | 16.74 | 12.36 | 7.33 | 14.89 | 11.63 | 10.41 | 7.03 | 7.23 | 9.50 | 8.43 | 11.51 | 11.08 |
| PL | 13.59 | 11.90 | 15.81 | 7.35 | 16.58 | 12.23 | 7.44 | 14.75 | 11.41 | 10.36 | 6.97 | 7.13 | 9.41 | 8.32 | 11.48 | 10.98 |
| FR-Online† | 13.70 | 11.98 | 16.01 | 7.40 | 16.73 | 12.31 | 7.38 | 14.88 | 11.61 | 10.40 | 7.04 | 7.26 | 9.45 | 8.43 | 11.47 | 11.07 |
| TFA-Online† | 12.71 | 11.00 | 13.56 | 6.65 | 14.96 | 10.96 | 6.72 | 12.76 | 10.30 | 8.77 | 6.35 | 6.49 | 8.33 | 7.81 | 10.42 | 9.85 |
| TTT++-Online† | 13.41 | 11.16 | 12.71 | 6.18 | 15.14 | 10.55 | 6.45 | 11.87 | 10.32 | 7.78 | 5.80 | **6.11** | 8.38 | 7.36 | 10.80 | 9.60 |
| EATA† | 13.32 | 11.79 | 15.51 | 7.35 | 16.41 | 12.07 | 7.19 | 14.47 | 11.37 | 10.09 | 6.96 | 7.10 | 9.30 | 8.34 | 11.22 | 10.83 |
| TENT† | 12.47 | 11.29 | 14.12 | 6.86 | 15.47 | 11.26 | 6.95 | 13.40 | 10.81 | 9.09 | 6.52 | 6.87 | 8.99 | 7.93 | 10.84 | 10.19 |
| **CAFA (Ours)** | **10.90** | **10.01** | **11.89** | 5.95 | **13.23** | **9.77** | **5.99** | **11.38** | **9.47** | **7.54** | 5.93 | 6.28 | 8.00 | **7.19** | **9.60** | **8.88** |
| Source | 71.80 | 63.00 | 65.65 | **26.15** | 73.00 | 46.43 | 31.03 | 48.51 | 48.84 | 35.35 | **25.45** | 28.77 | 30.76 | 38.34 | 43.74 | 45.12 |
| BN | 41.70 | 38.35 | 43.93 | 29.59 | 44.93 | 37.70 | 29.89 | 42.60 | 37.45 | 37.00 | 29.72 | 30.52 | 34.24 | 31.80 | 35.40 | 36.32 |
| PL | 41.26 | 38.08 | 43.45 | 29.59 | 44.68 | 37.35 | 29.69 | 42.33 | 37.33 | 36.50 | 29.55 | 30.21 | 33.80 | 31.58 | 35.11 | 36.03 |
| FR-Online† | 41.70 | 38.36 | 43.94 | 29.57 | 44.88 | 37.67 | 29.91 | 42.62 | 37.46 | 37.04 | 29.72 | 30.50 | 34.23 | 31.81 | 35.42 | 36.32 |
| TFA-Online† | 39.27 | 35.77 | 40.52 | 27.98 | 42.05 | 34.99 | 28.51 | 38.72 | 35.01 | 32.90 | 27.83 | 28.81 | 31.43 | 30.60 | 34.38 | 33.92 |
| TTT++-Online† | 38.39 | 34.51 | 38.95 | 26.52 | 41.53 | 33.65 | 27.64 | 37.75 | 34.06 | 31.32 | 26.78 | 27.22 | **30.18** | 29.46 | 33.26 | 32.75 |
| EATA† | 40.00 | 36.57 | 40.94 | 28.82 | 42.28 | 35.85 | 28.96 | 40.43 | 36.06 | 34.53 | 29.06 | 29.44 | 32.90 | 30.56 | 34.25 | 34.71 |
| TENT† | 38.37 | 34.72 | 39.50 | 27.41 | 40.91 | 33.81 | 27.99 | 39.10 | 34.69 | 31.97 | 27.36 | 27.94 | 31.80 | 29.09 | 32.94 | 33.17 |
| **CAFA (Ours)** | **36.97** | **33.53** | **37.41** | 26.21 | **38.96** | **32.46** | **27.26** | **36.82** | **33.64** | **30.13** | 26.53 | **27.07** | 30.20 | **28.23** | **32.05** | **31.83** |

Table 5. Classification error (%) on the CIFAR10-C (upper group) and CIFAR100-C (lower group) datasets with severity level 3 corruptions. † denotes the results obtained from the official codes.

| Method | Gaus. | Shot | Impu. | Defo. | Glas. | Moti. | Zoom | Snow | Fros. | Fog | Brig. | Cont. | Elas. | Pixe. | Jpeg. | Average |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Source | 24.43 | 14.91 | 18.87 | **5.32** | 37.16 | 11.58 | 6.75 | 18.33 | 11.29 | 6.76 | **5.38** | 5.94 | **7.06** | 10.35 | 14.42 | 13.24 |
| BN | 11.24 | 8.92 | 12.66 | 7.13 | 16.28 | 10.17 | 7.08 | 12.77 | 9.45 | 8.43 | 7.02 | 7.14 | 9.50 | 8.35 | 10.71 | 9.79 |
| PL | 11.17 | 8.81 | 12.71 | 6.92 | 16.27 | 10.10 | 7.15 | 12.67 | 9.43 | 8.41 | 7.02 | 6.96 | 9.45 | 8.31 | 10.66 | 9.74 |
| FR-Online† | 11.21 | 8.91 | 12.61 | 7.11 | 16.26 | 10.18 | 7.10 | 12.77 | 9.46 | 8.41 | 7.04 | 7.14 | 9.49 | 8.35 | 10.66 | 9.78 |
| TFA-Online† | 10.29 | 8.09 | 10.85 | 6.35 | 14.87 | 8.97 | 6.43 | 11.29 | 8.27 | 7.47 | 6.21 | 6.40 | 8.39 | 7.42 | 9.61 | 8.73 |
| TTT++-Online† | 10.47 | 7.96 | 10.16 | 5.89 | 14.92 | 8.63 | 6.26 | 10.40 | 8.26 | 6.80 | 5.78 | **5.86** | 7.84 | 7.07 | 9.75 | 8.40 |
| EATA† | 11.12 | 8.83 | 12.36 | 7.00 | 15.86 | 9.83 | 7.01 | 12.52 | 9.29 | 8.30 | 6.91 | 7.01 | 9.25 | 8.27 | 10.38 | 9.60 |
| TENT† | 10.52 | 8.15 | 11.45 | 6.65 | 15.03 | 9.64 | 6.67 | 11.18 | 8.69 | 7.44 | 6.50 | 6.72 | 8.80 | 7.82 | 9.85 | 9.01 |
| **CAFA (Ours)** | **9.22** | **7.51** | **9.65** | 5.78 | **13.22** | **8.20** | **5.86** | **9.71** | **7.57** | **6.55** | 5.87 | 6.02 | 7.64 | **6.95** | **8.90** | **7.91** |
| Source | 59.17 | 44.40 | 54.99 | **24.91** | 73.55 | 37.45 | 28.70 | 49.25 | 37.83 | 29.75 | **25.01** | 26.84 | **29.35** | 34.25 | 41.17 | 39.77 |
| BN | 37.60 | 33.23 | 39.55 | 29.11 | 44.32 | 33.91 | 29.63 | 38.89 | 33.89 | 33.74 | 29.05 | 29.74 | 33.21 | 30.94 | 34.03 | 34.06 |
| PL | 37.40 | 32.92 | 39.01 | 28.88 | 43.91 | 33.56 | 29.45 | 38.67 | 33.68 | 33.42 | 28.86 | 29.48 | 32.90 | 30.71 | 33.82 | 33.78 |
| FR-Online† | 37.61 | 33.23 | 39.56 | 29.11 | 44.34 | 33.91 | 29.62 | 38.89 | 33.94 | 33.69 | 29.05 | 29.75 | 33.21 | 30.91 | 34.05 | 34.06 |
| TFA-Online† | 35.27 | 31.10 | 36.71 | 27.86 | 42.04 | 32.00 | 28.65 | 36.03 | 31.18 | 30.22 | 27.60 | 28.32 | 30.65 | 29.30 | 33.11 | 32.00 |
| TTT++-Online† | 33.60 | 29.74 | 34.88 | 25.90 | 41.21 | 30.64 | 26.97 | 34.86 | 30.17 | 28.89 | 26.19 | 26.68 | 29.57 | 28.29 | 31.84 | 30.63 |
| EATA† | 36.01 | 32.23 | 37.20 | 28.17 | 41.81 | 32.72 | 28.73 | 37.01 | 32.60 | 32.17 | 28.13 | 28.88 | 31.80 | 29.69 | 32.92 | 32.67 |
| TENT† | 34.29 | 30.72 | 35.37 | 27.22 | 40.35 | 31.03 | 27.54 | 35.55 | 31.05 | 29.80 | 26.84 | 27.60 | 30.54 | 28.43 | 31.82 | 31.21 |
| **CAFA (Ours)** | **33.02** | **29.38** | **33.75** | 26.09 | **38.34** | **29.63** | **26.73** | **33.47** | **29.79** | **28.18** | 25.94 | **26.45** | 29.58 | **27.64** | **30.95** | **29.93** |

Table 6. Classification error (%) on the CIFAR10-C (upper group) and CIFAR100-C (lower group) datasets with severity level 2 corruptions. † denotes the results obtained from the official codes.

| Method | Gaus. | Shot | Impu. | Defo. | Glas. | Moti. | Zoom | Snow | Fros. | Fog | Brig. | Cont. | Elas. | Pixe. | Jpeg. | Average |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Source | 14.00 | 10.09 | 11.62 | **5.24** | 38.69 | 7.75 | 7.40 | 9.48 | 8.14 | **5.74** | **5.37** | **5.47** | **7.69** | 6.69 | 10.51 | 10.26 |
| BN | 8.89 | 8.08 | 10.17 | 7.04 | 16.16 | 8.55 | 7.55 | 9.44 | 7.66 | 7.47 | 7.00 | 6.92 | 9.81 | 7.50 | 8.76 | 8.73 |
| PL | 8.83 | 7.90 | 10.17 | 6.94 | 16.05 | 8.42 | 7.59 | 9.40 | 7.54 | 7.41 | 6.93 | 6.81 | 9.77 | 7.38 | 8.70 | 8.66 |
| FR-Online† | 8.92 | 8.08 | 10.19 | 7.03 | 16.12 | 8.57 | 7.53 | 9.48 | 7.67 | 7.49 | 7.00 | 6.92 | 9.80 | 7.47 | 8.77 | 8.74 |
| TFA-Online† | 7.94 | 7.24 | 8.95 | 6.26 | 14.38 | 7.63 | 7.11 | 8.37 | 7.10 | 6.45 | 6.23 | 6.02 | 8.64 | 6.75 | 7.69 | 7.79 |
| TTT++-Online† | 8.23 | 6.97 | 8.52 | 5.61 | 15.17 | 7.43 | 6.84 | **7.66** | 6.55 | 5.95 | 5.61 | 5.80 | 8.39 | 6.65 | 7.86 | 7.55 |
| EATA† | 8.81 | 7.86 | 10.12 | 6.99 | 15.70 | 8.43 | 7.54 | 9.31 | 7.52 | 7.40 | 6.91 | 6.81 | 9.60 | 7.38 | 8.61 | 8.60 |
| TENT† | 8.25 | 7.50 | 9.71 | 6.54 | 14.75 | 7.94 | 7.15 | 8.71 | 6.94 | 6.62 | 6.43 | 6.48 | 9.35 | 7.29 | 8.37 | 8.14 |
| **CAFA (Ours)** | **7.37** | **6.54** | **8.36** | 5.80 | **13.07** | **6.99** | **6.34** | 7.84 | **6.38** | 6.04 | 5.70 | 5.76 | 8.03 | **6.35** | **7.44** | **7.20** |
| Source | 42.81 | 35.37 | 40.39 | **24.70** | 74.15 | 31.29 | 29.80 | 35.16 | 31.07 | **25.20** | **24.79** | **25.23** | **30.04** | 28.64 | 34.83 | 34.23 |
| BN | 32.74 | 30.64 | 34.21 | 28.69 | 44.57 | 32.11 | 30.34 | 33.10 | 30.59 | 29.69 | 28.55 | 28.90 | 34.23 | 29.94 | 31.52 | 31.99 |
| PL | 32.61 | 30.50 | 33.70 | 28.52 | 44.10 | 31.83 | 30.19 | 32.82 | 30.52 | 29.25 | 28.46 | 28.61 | 33.84 | 29.72 | 31.27 | 31.73 |
| FR-Online† | 32.74 | 30.67 | 34.22 | 28.71 | 44.58 | 32.15 | 30.35 | 33.08 | 30.62 | 29.68 | 28.58 | 28.92 | 34.21 | 29.95 | 31.53 | 32.00 |
| TFA-Online† | 30.64 | 28.78 | 31.88 | 27.43 | 41.80 | 30.58 | 29.04 | 30.70 | 29.27 | 27.84 | 27.31 | 27.56 | 31.11 | 28.58 | 30.29 | 30.19 |
| TTT++-Online† | 29.63 | 28.25 | 30.33 | 25.63 | 41.00 | 29.07 | 27.36 | 29.76 | 27.81 | 26.35 | 26.17 | 26.14 | 30.11 | 27.34 | 29.15 | 28.94 |
| EATA† | 31.73 | 29.93 | 32.78 | 27.82 | 42.00 | 31.03 | 29.45 | 31.99 | 29.76 | 28.40 | 27.89 | 28.01 | 32.85 | 29.01 | 30.53 | 30.88 |
| TENT† | 30.07 | 28.53 | 31.01 | 26.84 | 40.33 | 29.45 | 28.18 | 30.56 | 28.49 | 27.14 | 26.71 | 27.07 | 31.33 | 27.76 | 29.12 | 29.51 |
| **CAFA (Ours)** | **28.82** | **27.70** | **30.09** | 25.84 | **38.38** | **28.02** | **27.09** | **29.06** | **27.36** | 26.23 | 25.80 | 25.91 | 30.33 | **26.94** | **28.63** | **28.41** |

Table 7. Classification error (%) on the CIFAR10-C (upper group) and CIFAR100-C (lower group) datasets with severity level 1 corruptions. † denotes the results obtained from the official codes.

| Trial | Gaus. | Shot | Impu. | Defo. | Glas. | Moti. | Zoom | Snow | Fros. | Fog | Brig. | Cont. | Elas. | Pixe. | Jpeg. | Average |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 14.05 | 13.06 | 21.35 | 8.07 | 20.45 | 11.10 | 6.96 | 11.92 | 11.31 | 13.22 | 7.06 | 7.14 | 16.05 | 9.65 | 11.70 | 12.21 |
| 1 | 13.79 | 12.90 | 21.05 | 8.13 | 21.08 | 10.85 | 6.92 | 11.92 | 11.20 | 13.10 | 6.83 | 7.22 | 16.03 | 9.63 | 11.57 | 12.15 |
| 2 | 14.07 | 12.60 | 21.01 | 7.89 | 20.73 | 10.71 | 6.86 | 11.87 | 11.24 | 13.46 | 6.92 | 6.94 | 16.16 | 9.80 | 11.55 | 12.12 |
| 3 | 13.80 | 12.99 | 20.83 | 8.00 | 20.62 | 10.55 | 6.71 | 11.89 | 11.64 | 13.68 | 7.11 | 7.04 | 16.22 | 9.82 | 11.25 | 12.14 |
| 4 | 13.70 | 12.71 | 21.07 | 8.09 | 20.94 | 10.68 | 6.87 | 11.96 | 11.41 | 13.82 | 7.04 | 7.05 | 15.90 | 9.67 | 11.68 | 12.17 |
| std | 0.17 | 0.19 | 0.19 | 0.09 | 0.25 | 0.21 | 0.10 | 0.03 | 0.18 | 0.30 | 0.11 | 0.11 | 0.12 | 0.09 | 0.18 | 0.03 |

Table 8. Classification error (%) on the severity level 5 corruptions in the CIFAR10-C dataset with different random seeds.

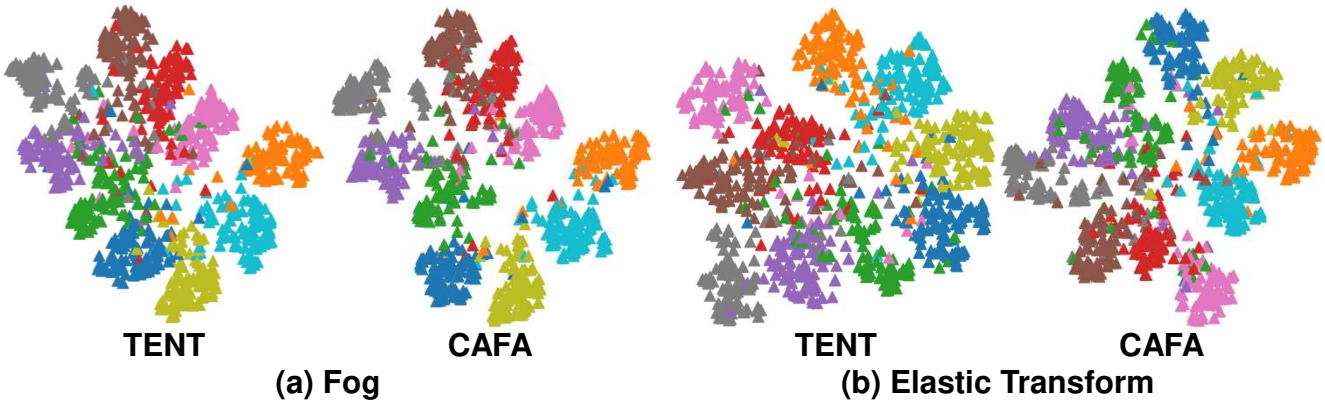


Figure 1. t-SNE visualizations of ours (CAFA) and TENT from (a) Fog and (b) Elastic Transform corruptions with severity level 5 in the CIFAR10-C dataset.

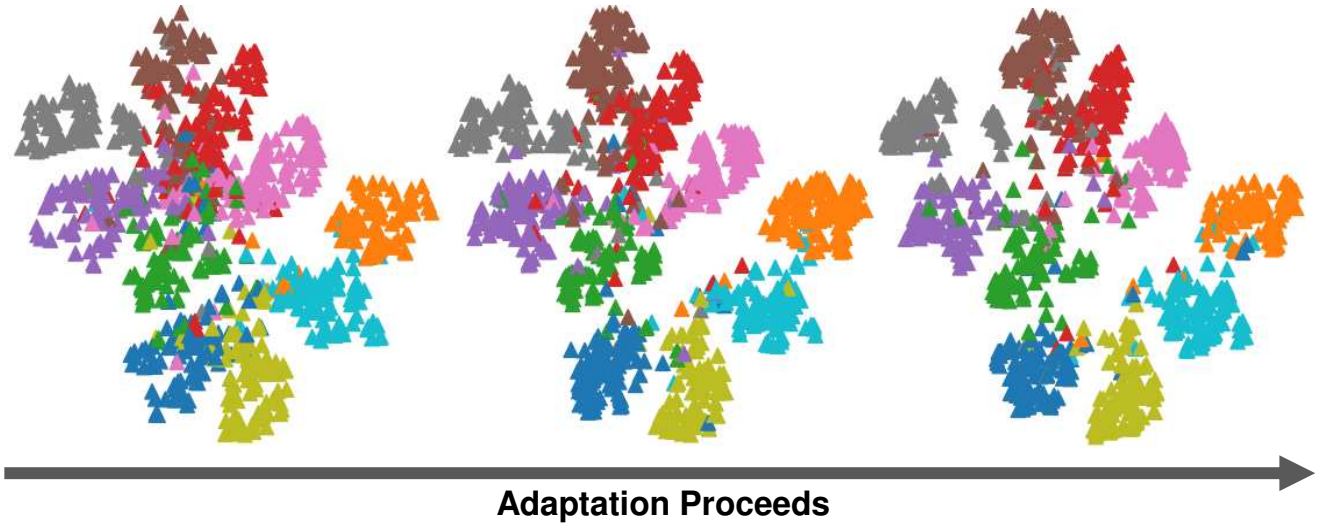


Figure 2. Change of the representation space of test samples from our method as adaptation proceeds. Representation space is visualized by the t-SNE algorithm, and visualizations are obtained from the Fog corruption with severity level 5 in the CIFAR10-C dataset.

## References


[1] Dian Chen, Dequan Wang, Trevor Darrell, and Sayna Ebrahimi. Contrastive test-time adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 295–305, 2022. 1

[2] Cian Eastwood, Ian Mason, Chris Williams, and Bernhard Schölkopf. Source-free adaptation to measurement shift via bottom-up feature restoration. In *International Conference on Learning Representations*, 2021. 1

[3] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *International conference on machine learning*, pages 448–456. PMLR, 2015. 1

[4] Yusuke Iwasawa and Yutaka Matsuo. Test-time classifier adjustment module for model-agnostic domain generalization. *Advances in Neural Information Processing Systems*, 34, 2021. 1

[5] Dong-Hyun Lee. Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks. In *Workshop on challenges in representation learning, ICML*, volume 3, page 896, 2013. 1

[6] Yuejiang Liu, Parth Kothari, Bastien van Delft, Baptiste Bellot-Gurlet, Taylor Mordan, and Alexandre Alahi. Ttt++: When does self-supervised test-time training fail or thrive? *Advances in Neural Information Processing Systems*, 34, 2021. 1

[7] Zachary Nado, Shreyas Padhy, D Sculley, Alexander D'Amour, Balaji Lakshminarayanan, and Jasper Snoek. Evaluating prediction-time batch normalization for robustness under covariate shift. *arXiv preprint arXiv:2006.10963*, 2020. 1

[8] 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*, pages 16888–16905. PMLR, 2022. 1

[9] Yu Sun, Xiaolong Wang, Zhuang Liu, John Miller, Alexei Efros, and Moritz Hardt. Test-time training with self-supervision for generalization under distribution shifts. In Hal Daumé III and Aarti Singh, editors, *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pages 9229–9248. PMLR, 13–18 Jul 2020. 1

[10] Dequan Wang, Evan Shelhamer, Shaoteng Liu, Bruno Olshausen, and Trevor Darrell. Tent: Fully test-time adaptation by entropy minimization. In *International Conference on Learning Representations*, 2021. 1, 2

[11] Marvin Mengxin Zhang, Sergey Levine, and Chelsea Finn. MEMO: Test time robustness via adaptation and augmentation. In *NeurIPS 2021 Workshop on Distribution Shifts: Connecting Methods and Applications*, 2021. 1