Supplementary

A. Peak Detection Algorithm

Here, we provide details of the Peak Detection Algorithm.

Algorithm 2 Online Peak Detection

Require: window_size: Size of sliding window
Require: threshold: Z-score threshold for peak detection
Require: k: Influence value for the running statistics
Require: data: Incoming $\bar{\alpha}^t$ stream
  for each incoming $\bar{\alpha}^t$ **do**
    if sliding window length $\leq$ window_size **then**
      Append $\bar{\alpha}^t$ to window
      if sliding window length $=$ window_size **then**
        Initialize running μ_r and σ_r as simple mean and standard deviation within sliding window
      end if
    else
      Calculate z-score for $\bar{\alpha}^t$: $z = \frac{\bar{\alpha}^t - \mu_r}{\sigma_r}$
      if $z >$ threshold **then**
        Consider $\bar{\alpha}^t$ as a peak or anomaly
      end if
      Slide the window by removing the oldest data point and appending $\bar{\alpha}^t$
      Calculate current mean μ_c and std σ_c of the sliding window
      Update the running μ_r and σ_r using the current μ_c and σ_c with following eqns:
      $\mu_r = (1 - k) \times \mu_r + k \times \mu_c$
      $\sigma_r = (1 - k) \times \sigma_r + k \times \sigma_c$
    end if
  end for

B. Additional $\bar{\alpha}^t$ Plots

In this section, we present additional visualizations of the $\bar{\alpha}^t$ values obtained from our experiments. We first include visualizations of the $\bar{\alpha}^t$ with respect to incoming test batches for the first cycle of domains from the experiments of digit in Fig. 1 and office-home in Fig. 2. We also show the plot for ImageNet-C with severity level 5 in Fig. 3.

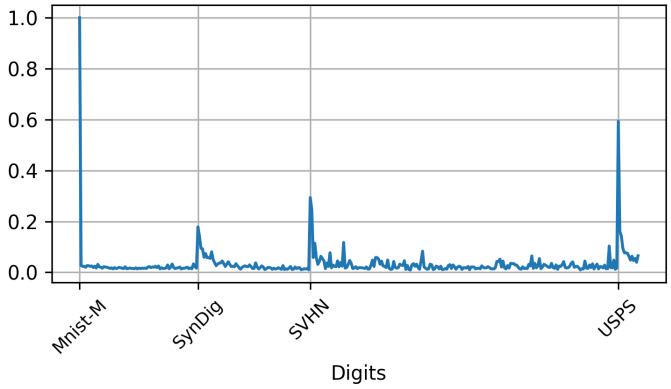

Figure 1. $\bar{\alpha}^t$ plot with respect to incoming test batches. The source model is trained on MNIST dataset and during test-time sequentially adapted to the remaining digit datasets.

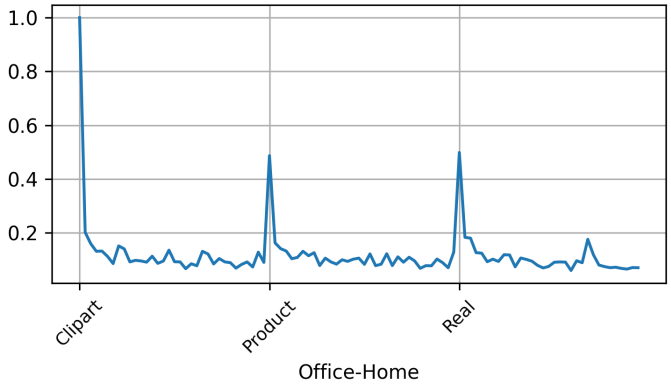

Figure 2. $\bar{\alpha}^t$ plot with respect to incoming test batches. The source model is trained on Art dataset and during test-time sequentially adapted to the remaining office-home datasets.

C. Decision Diagram

In this section, we visualize the domain change locations detected by our algorithm.

Fig. 4 shows the visualization for CIFAR-100C. It can be observed that our method is capable of correctly detecting domain changes most of the time. It should be noted that, as our method rapidly aligns the running BN stats with the

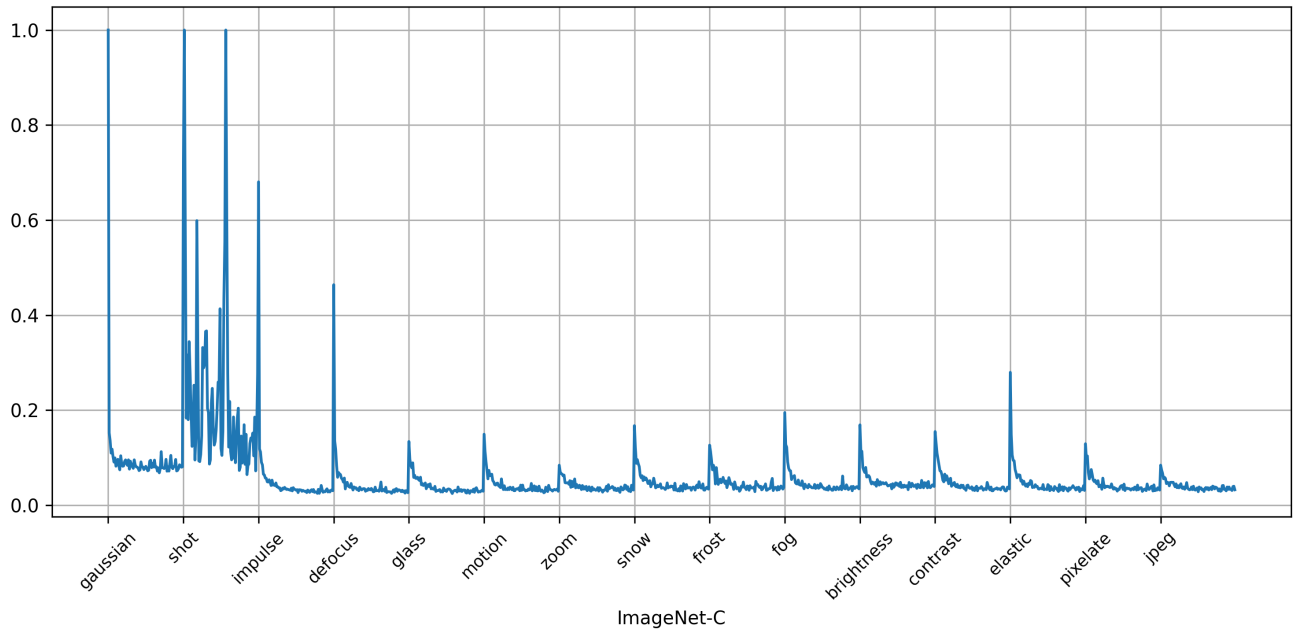

Figure 3. $\bar{\alpha}^t$ plot with respect to incoming test batches for ImageNet-C dataset

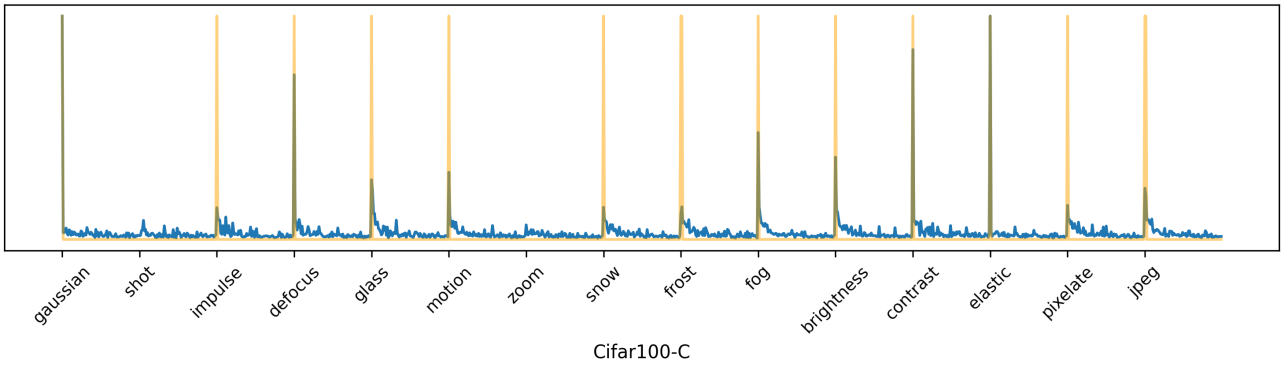

Figure 4. Our algorithm's decision points are overlaid onto $\bar{\alpha}^t$ plot with respect to incoming test batches for CIFAR100-C dataset. Instances where our proposed algorithm detects a domain change are marked by orange lines.

new domain, the false positive rates are also extremely low.

D. Sensitivity Analysis

The critical hyperparameter of the peak detection algorithm is the threshold value. We have used a value of 15 standard deviations for all our experiment. Here, we change the threshold value and plot the corresponding decision diagrams for CIFAR-100C.

From Fig. 5, it can be observed that our method is less sensitive to different values of the threshold and yields same classification performance. However, excessively high values could potentially overlook peak values, as observed by the last subplot of Fig. 5.

E. Limitations

If two domains share very similar features, i.e., domain gap is very low, our method's ability to detect domain changes can be diminished, as the global BN statistics between the two domains will be similar. However, it should also be noted that when the domain gap is low, the problem of forgetting and error accumulation are very much reduced, hence the need for taking additional measure to address the problems remains less critical. To demonstrate this, we perform the same classification experiment as Table 1 of main paper on CIFAR-100C, but with a severity level 1. Such low severity level results in images of different domains to be almost indistinguishable to the human eye. The results are shown in Table 1.

Table 1. Classification error rate on CIFAR-100C with a severity level of 1. When the domain gap is low, the continual method's performance is on par with the online model. Nonetheless, our method managed to improve the continual method's performance even more.

Method	GN	SN	IN	DB	GB	MB	ZB	Snow	Frost	Fog	Bright	Contrast	Elastic	Pixel	JPEG	Mean
Tent Online	26.2	24.1	22.8	22.2	30.7	23.5	22.9	23.6	24.1	22.1	21.9	22.6	26.7	23.7	28.7	24.4
Tent Continual	26.2	24.9	24.6	24.1	31.8	26.5	26.2	26.7	26.7	25.9	25.8	25.9	29.2	26.2	30.1	26.7
Tent+Ours	26.4	25.2	22.9	22.5	30.6	23.5	23.2	23.7	24.9	24.0	23.9	23.8	28.5	25.2	30.4	25.2

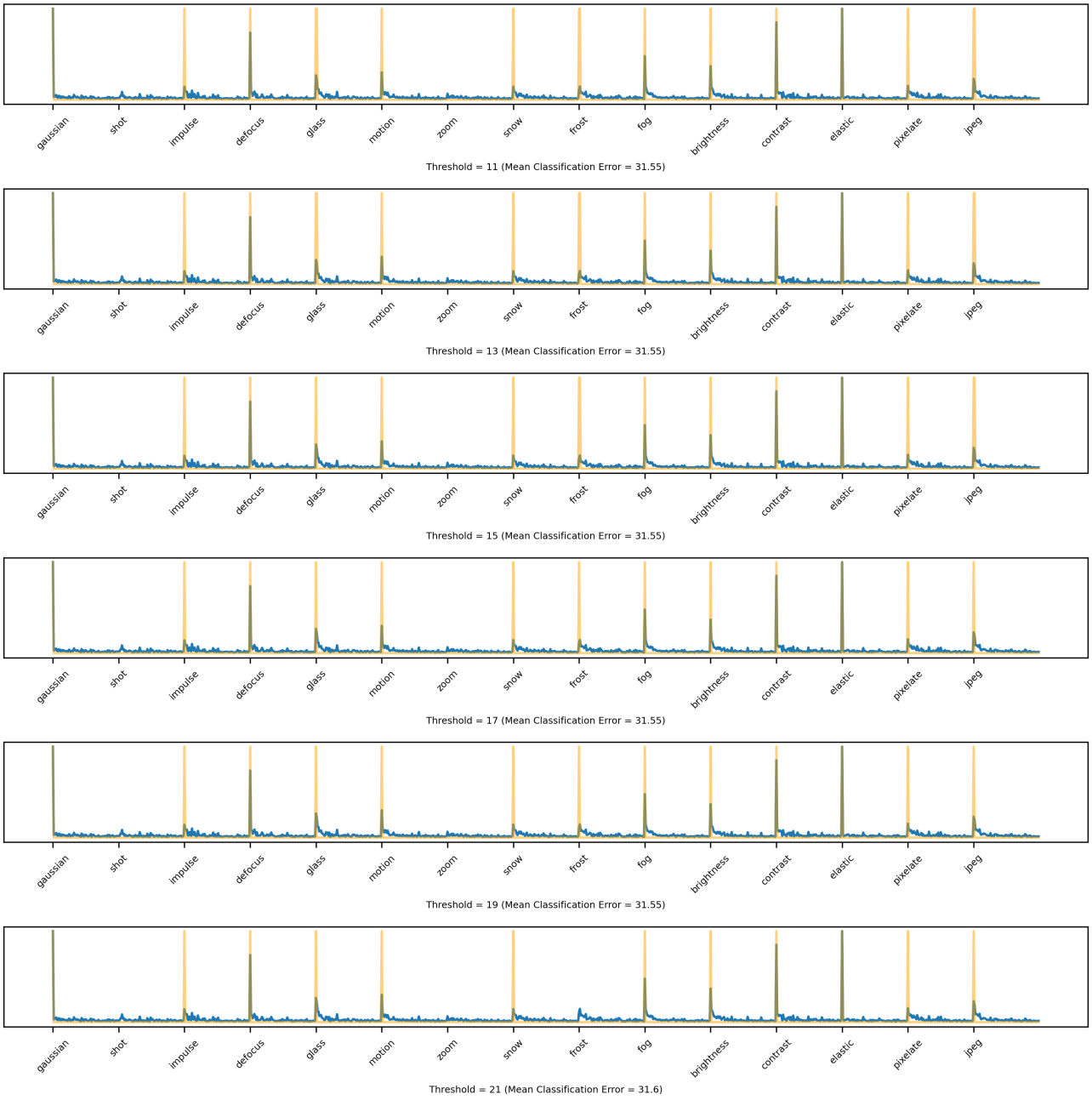

Figure 5. Decision diagram corresponding to different threshold values of online peak detection algorithm on CIFAR-100C. It can be observed that our method is less sensitive to different values of the threshold and yields same classification performance.

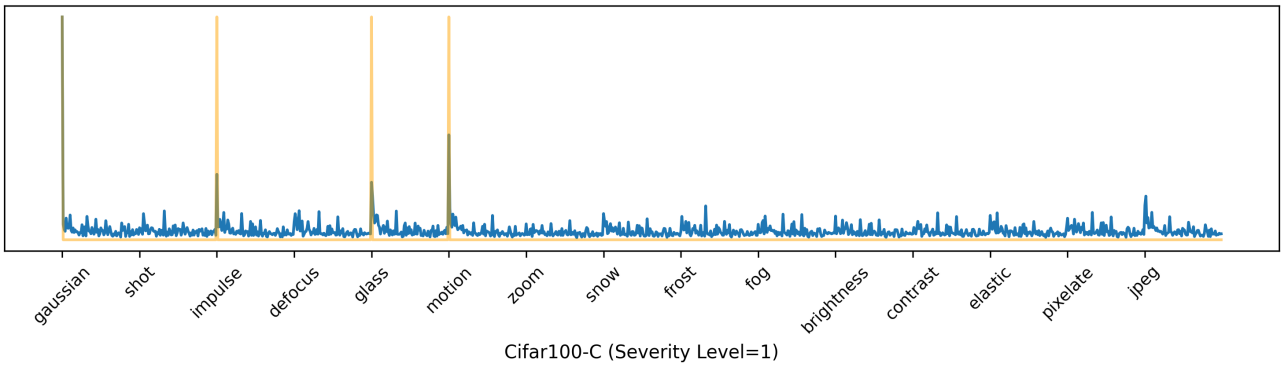

Figure 6. Decision diagram corresponding to CIFAR100-C with severity level 1.

From the table it can be observed that Tent-Continual exhibits only a marginal performance decline of 2.3% compared to Tent-Online. This is in contrast to the experiment in Table 1 of the main paper with a severity level of 5, which led to a performance drop of 37.1%. Clearly, in scenarios with minimal domain gap, the challenges of error accumulation and forgetting are significantly mitigated. Furthermore, it's worth highlighting that even in such a challenging scenario of low domain gap, our method has managed to enhance performance when integrated with Tent.

Additionally, in Figure 6, we have visualized the decision thresholds for the severity 1 case. The figure highlights that the absence of a peak in BN statistics is noticeable when encountering new domains, primarily due to the similarity of global BN statistics across domains. Despite this challenging situation, our method successfully identifies several domain changes, leading to performance enhancement over the continual model, as demonstrated in Table 1.