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 $\tilde{\alpha}_t$ stream

for each incoming $\tilde{\alpha}_t$ do  
if sliding window length $\leq$ window size then  
Append $\tilde{\alpha}_t$ to window  
if sliding window length $=$ window size then  
Initialize running $\mu_r$ and $\sigma_r$ as simple mean and standard deviation within sliding window  
end if  
else  
Calculate z-score for $\tilde{\alpha}_t$: $z = \frac{\tilde{\alpha}_t - \mu_r}{\sigma_r}$  
if $z >$ threshold then  
Consider $\tilde{\alpha}_t$ as a peak or anomaly  
end if  
end if  
end for

B. Additional $\tilde{\alpha}_t$ Plots

In this section, we present additional visualizations of the $\tilde{\alpha}_t$ values obtained from our experiments. We first include visualizations of the $\tilde{\alpha}_t$ 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.

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
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.

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

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.
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.