A SUPPLEMENTARY MATERIAL

A.1 BREAKDOWN OF SSL METHODS

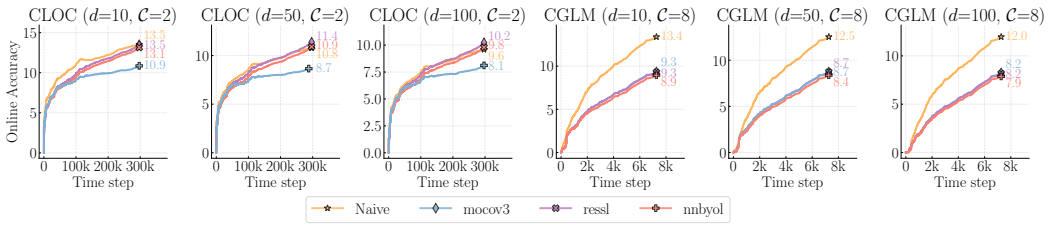

Figure 6: Comparison of the best performing SSL based methods after hyper-parameter tuning

In Figure 6 we show the performance of the best performing SSL based methods after hyper-parameter tuning. We observe that the performance of the SSL methods is highly dependent on the dataset and the delay setting. However, we apart from MoCo v3 (Chen et al., 2021), the methods perform similarly to Naive on CLOC. On the other hand on CGLM they have insignificant differences in performance, but consistently underperform Naive.

A.2 BREAKDOWN OF TTA METHODS

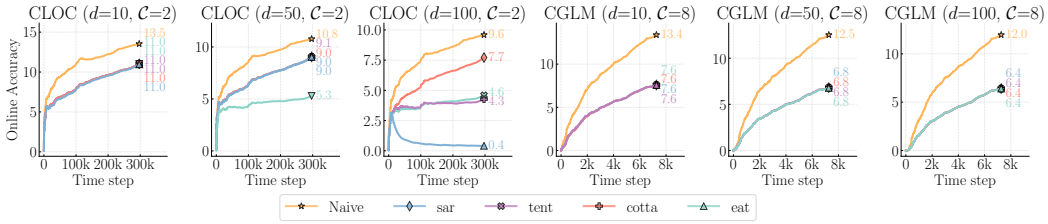

Figure 7: Comparison of the best performing TTA based methods after hyper-parameter tuning

In Figure 7 we show the performance of the best performing TTA based methods after hyper-parameter tuning. We observe that the performance of the TTA methods are consistently worse than Naive on both CLOC and CGLM, under all delay settings. We observe that in the most severe delay scenario ($d = 100$) the performance of EAT (Niu et al., 2022) and SAR (Niu et al., 2023) is comparable to Naive on CLOC, while CoTTA (Wang et al., 2022a) avoids the catastrophic performance drop.

A.3 COMPARISON OF SSL BASED METHODS TO NAIVE WHEN USING SAME AMOUNT OF SUPERVISED DATA

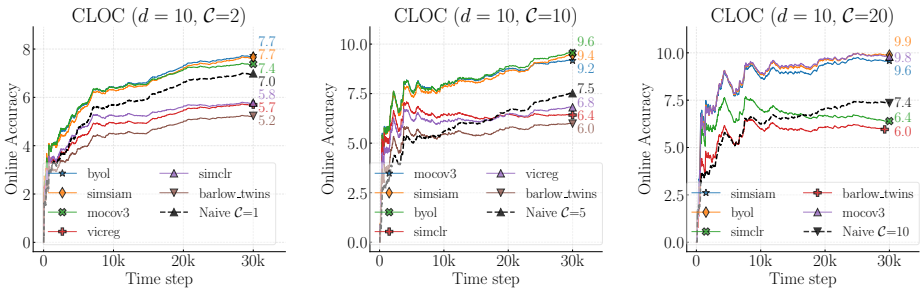

Figure 8: Detailed breakdown of various SSL methods from each family. Results are shown across varying number of parameter updates $\mathcal{C} = 2, 10, 20$ under the $d = 10$ scenario.

In Figure 8, we show that when trained on equal amount of supervised data, SSL based methods perform outperform Naive, however the performance gap is not as significant as in the case of using the same computational budget, as shown in the main Figure 4.

A.4 Examples of the Importance Weighted Memory Sampling on CLOC

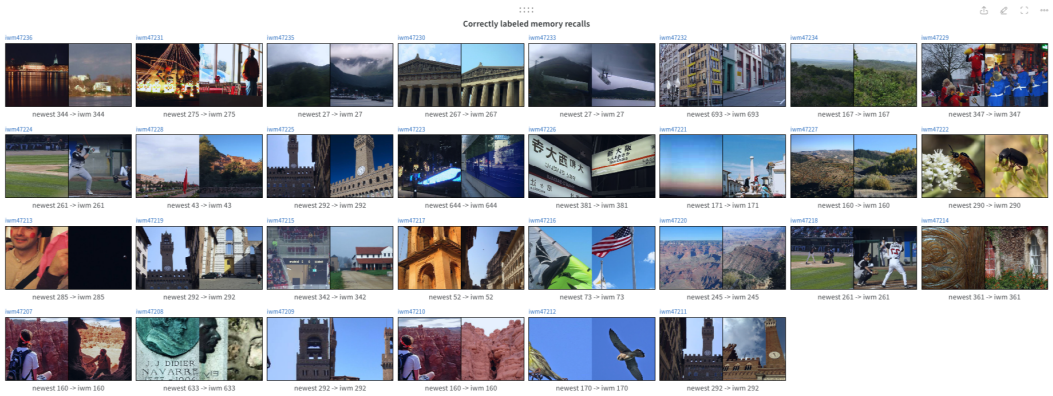

Figure 9: **Correctly labeled memory recalls.** In the subfigure's caption "Newest" refers to the newest unsupervised image observed by the model and "iwm" refers to the sample drawn from the memory by our proposed sampling method. The numbers refer to the corresponding true label IDs.

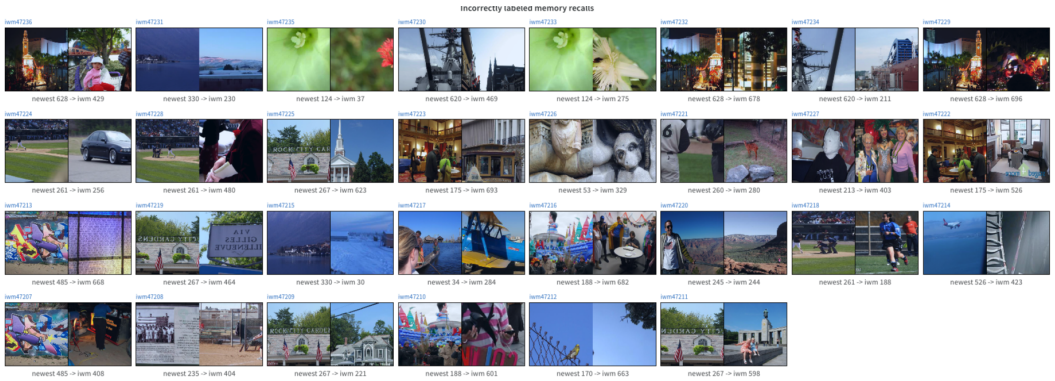

Figure 10: **Incorrectly labeled memory recalls.** In the subfigure's caption "Newest" refers to the newest unsupervised image observed by the model and "iwm" refers to the sample drawn from the memory by our proposed sampling method. The numbers refer to the corresponding true label IDs.

On CLOC, we report similar scores to Naive due to high noise in the data. To provide evidence for our claims we visualize the supervised data sampled from the memory buffer by our Importance Weighted Memory Sampling method. In Figure 9, we show that our method is capable of guessing the correct location of the unsupervised sample (the left hand side of the image pairs) and recalling a relevant sample from memory. In contrast, the incorrect memory recalls hurt the performance even though the content of the samples might match. We illustrate such cases in Figure 10, where it is obvious that in some cases the underlying image content has no information related to the location where the picture was taken at. In such scenarios, the only way a classifier can correctly predict the labels is by exploiting label correlations, *e.g.*classifying all close-up images of flowers to belong to the same geo-location, even though the flowers are not unique to the location itself. Or consider the pictures taken at social gatherings (second row, second column from the right), where a delayed classifier without being exposed to that specific series of images has no reason to correctly predict the location ID. Our claims are reinforced by the findings of Hammoud et al. (2023).