

A. Appendix

A.1. Character representation

In the results, we show that the model can obtain good results when training a linear probing to identify the character. this shows that the model has the information within the representation to solve the task. However, since we work in a continuous environment, it is important to know when the model obtains the relevant information to solve the task.

In Figure 5, we can observed that the model is capable of accumulating knowledge. Even though the performance without memory is worse, it is important to note that the model is capable of good performance, both for the IID and OOD tests.

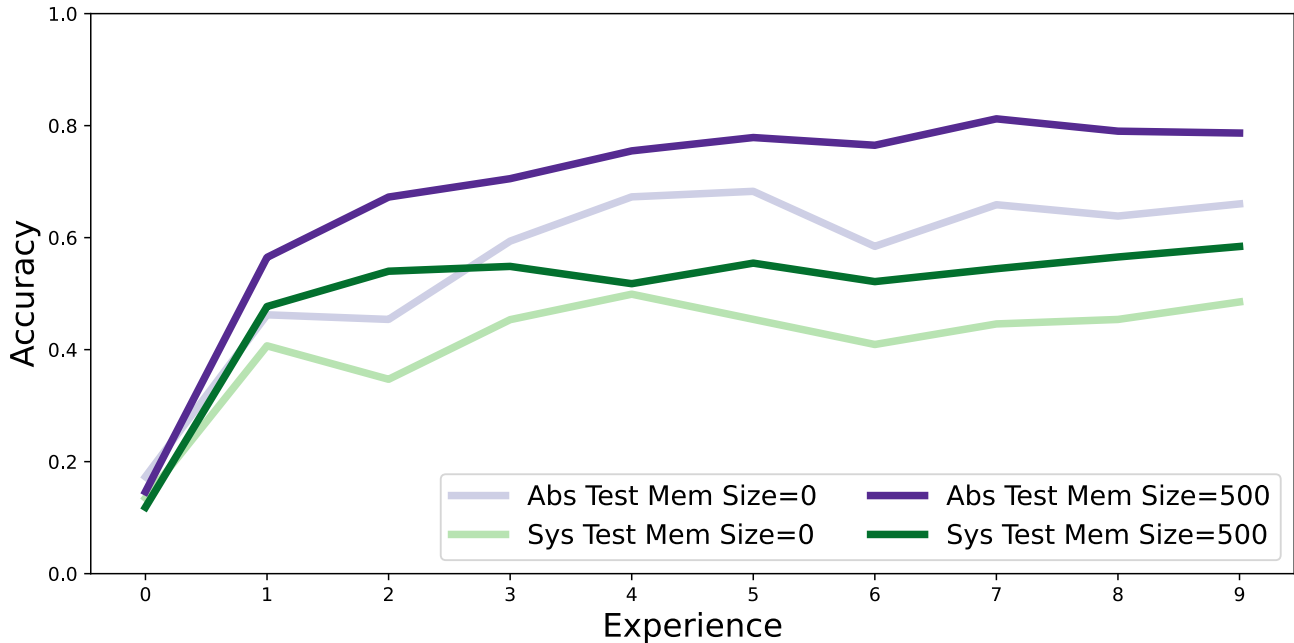


Figure 5. Character probing Accuracy after each experience for the 10 task scenario.

A.2. Probing Methodology

Using the representations at various stages of the model, we trained a linear probe to test the information stored in these. We used a linear two-layer perceptron, the SGD with momentum training algorithm. We used a grid-search over the hidden layer size, $\{16, 32, 64, 128, 256\}$, and learning rate $\{10^{-1}, 5 \cdot 10^{-2}, 10^{-2}, 5 \cdot 10^{-3}\}$, for model selection.

We use the probe to understand how much information exists in the model representation when trained in different scenarios. Figures 4(a) and 4(b) show a summary of the results. Here we show more clearly each of the results.

Figures 6 and 7 show the performance of applying the probe for the font task with the training distribution, IID, and the complete distribution, IID + OOD, respectively.

Figures 8 and 9 shows the performance of applying probing for the character task with the training distribution, IID, and the complete distribution, IID + OOD, respectively.

Figures 10 and 11 show the performance of applying probing for the character task with the training distribution, IID, and the complete distribution, IID + OOD, respectively.

Studying Generalization on Memory-Based Methods in Continual Learning

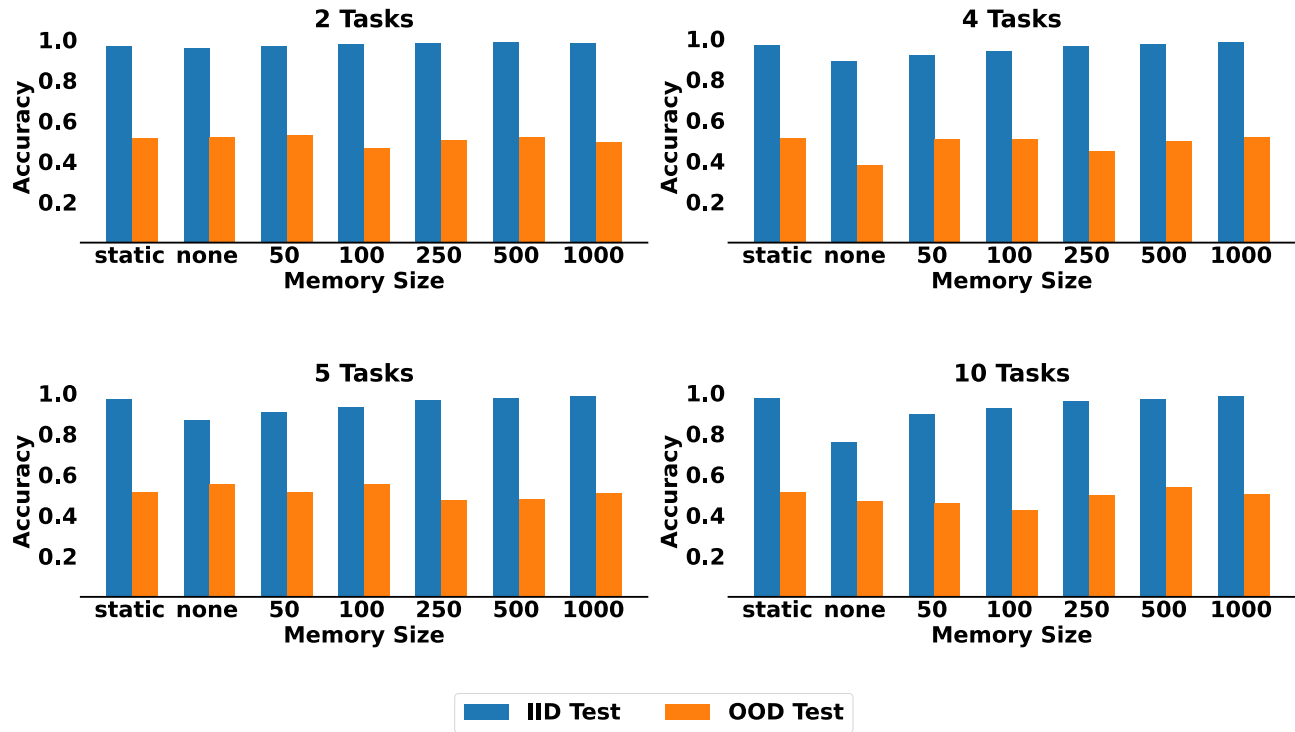


Figure 6. Result of the font probing task using the training data distribution, and testing on the IID (blue) and OOD (orange) test set.



Figure 7. Result of the font probing task using the complete data distribution, IID + OOD, and testing on the IID (blue) and OOD (orange) test set.

Studying Generalization on Memory-Based Methods in Continual Learning

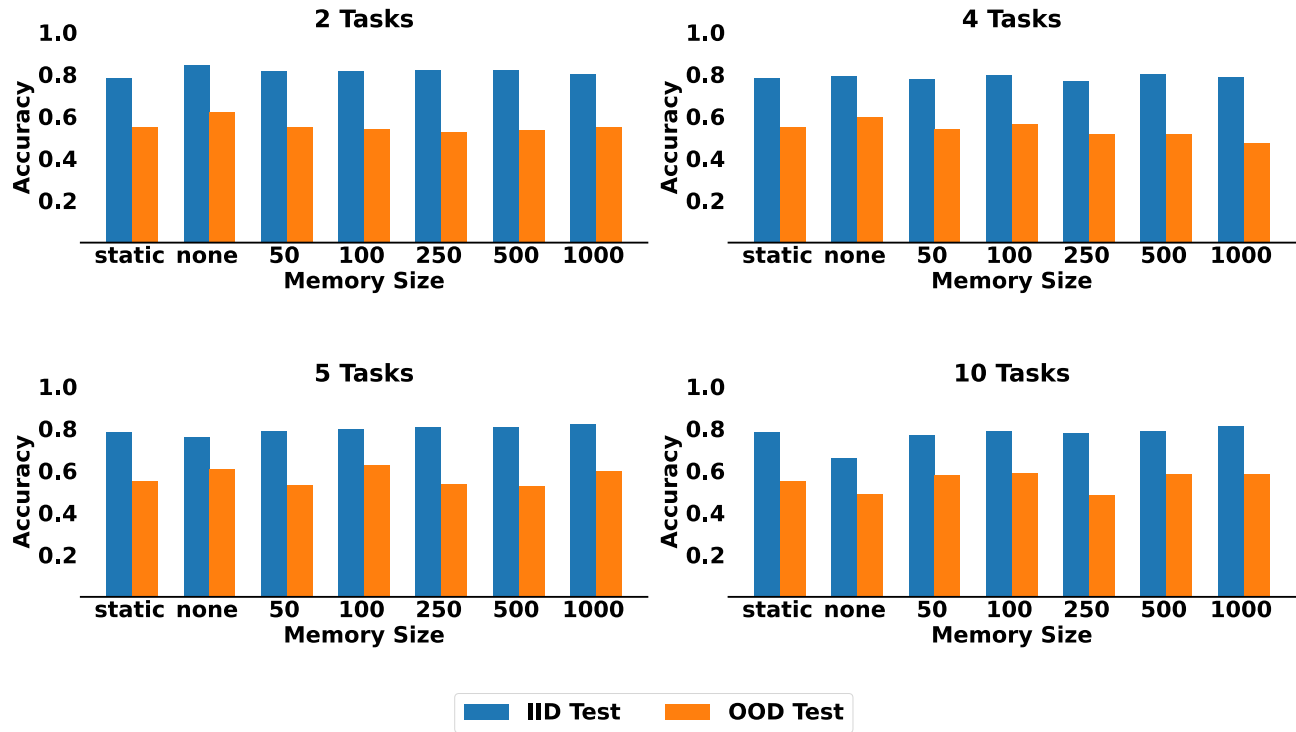


Figure 8. Result of the char probing task training with only the train distribution, IID. Testing on the IID (blue) and OOD (orange) test set.

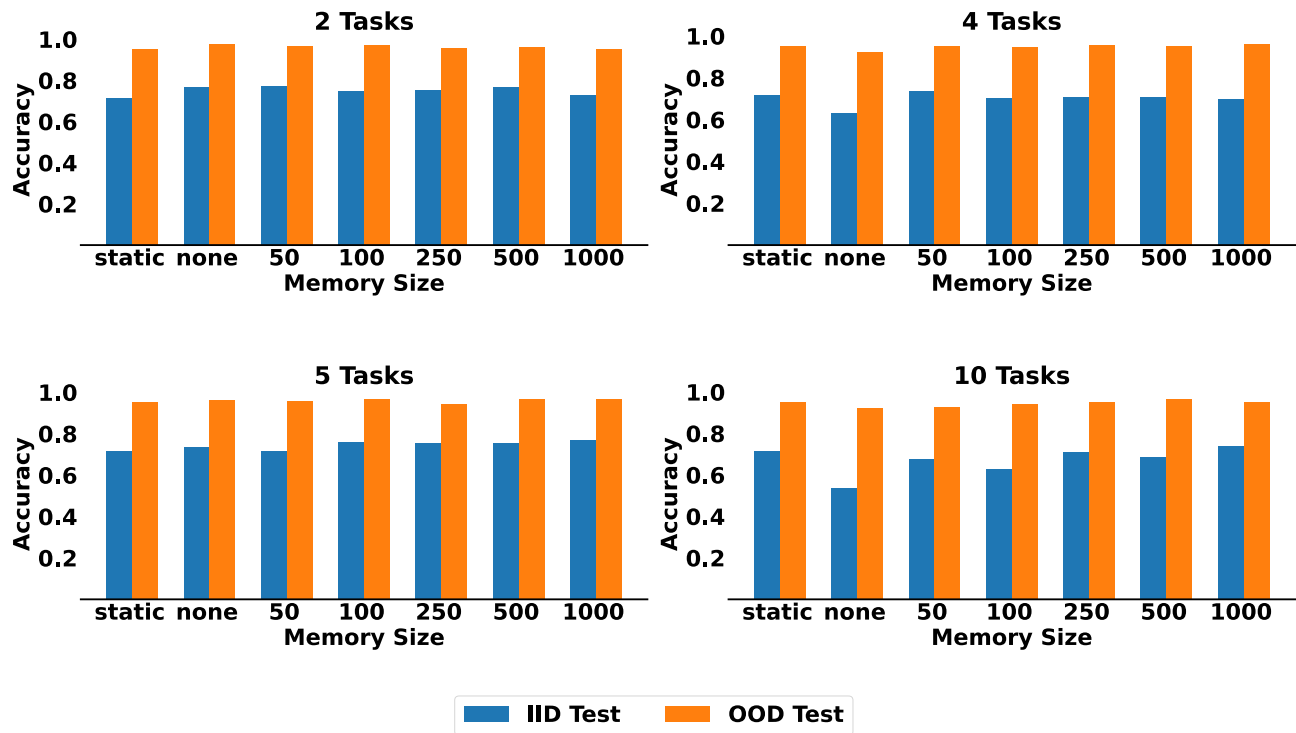


Figure 9. Result of the char probing task using the complete data distribution, IID + OOD. Testing on the IID (blue) and OOD (orange) test set.

Studying Generalization on Memory-Based Methods in Continual Learning

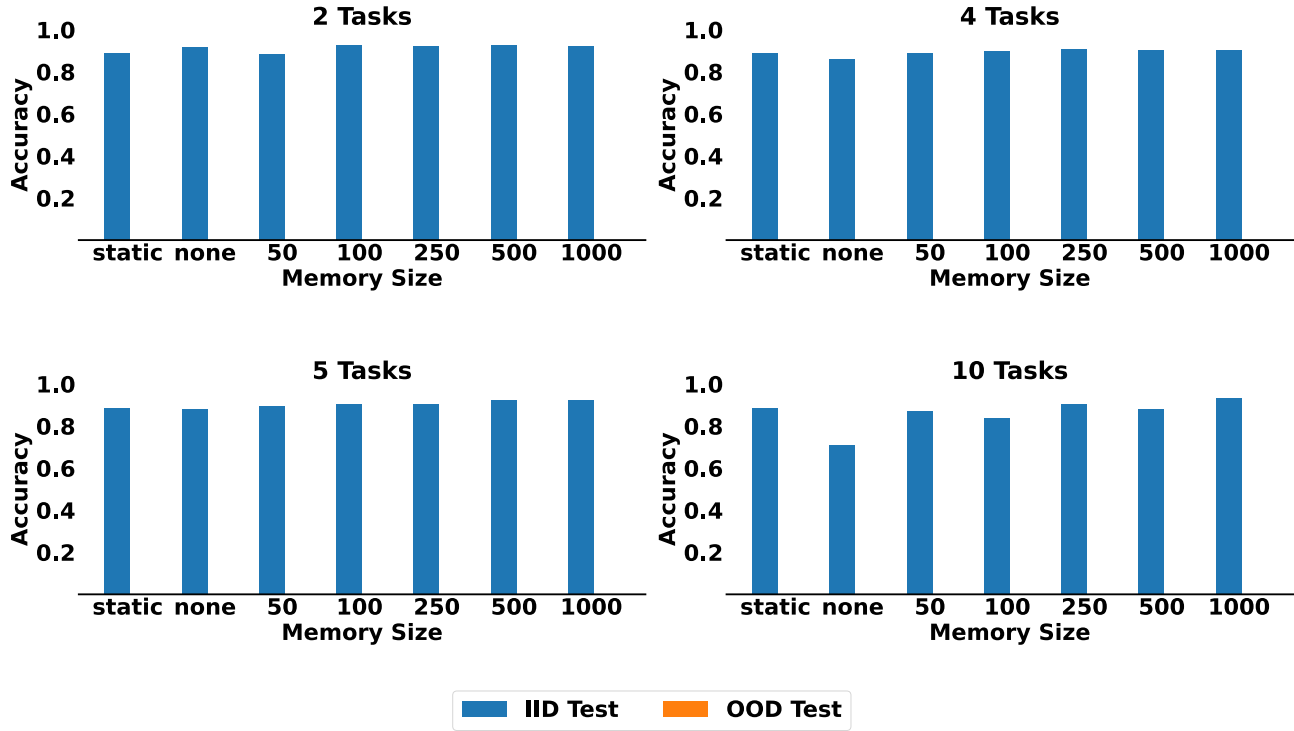


Figure 10. Result of the char/font probing task using the training set. Note that the result achieved in the OOD test are zero because the models haven't seen any data with that label during training. Similar to previous figures, IID is blue and OOD is orange.

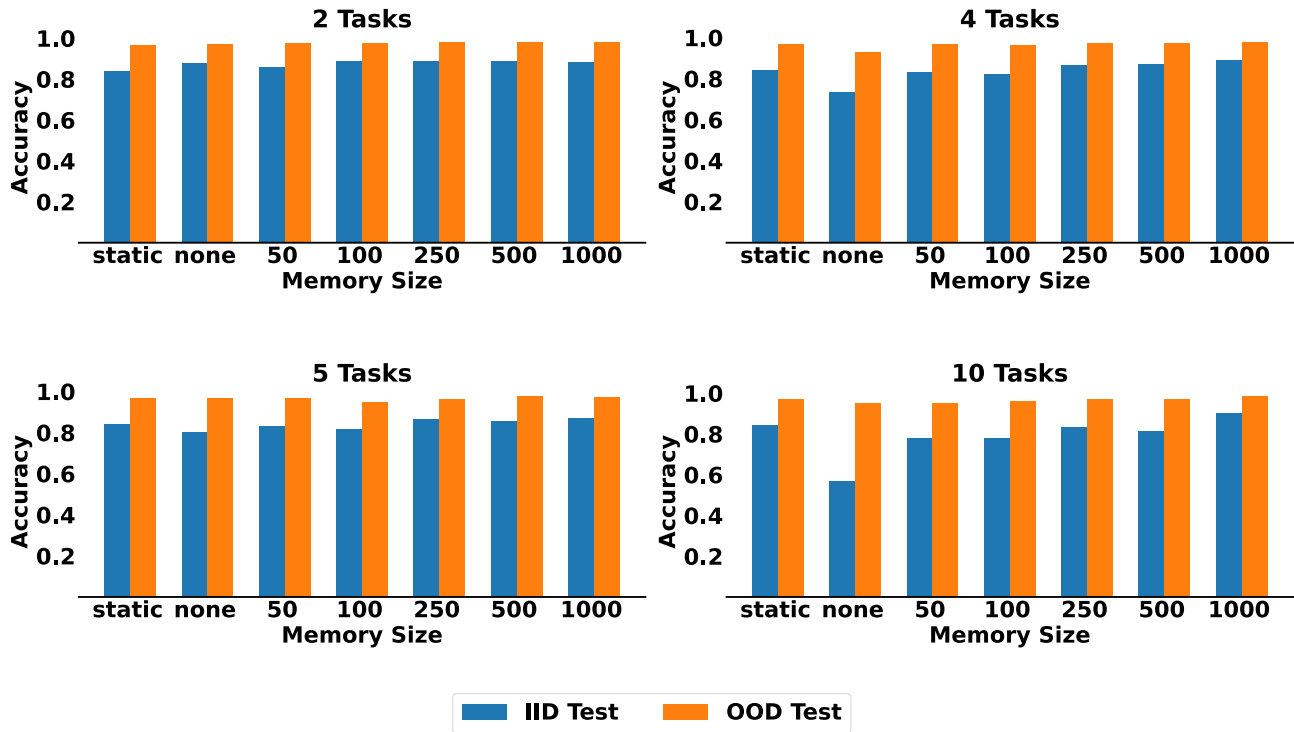


Figure 11. Result of the char/font probing task using the complete data distribution, IID + OOD. Testing on the IID (blue) and OOD (orange) test set.