Supplementary Materials

1 Experimental Setup

1.1 Expert Teacher Model

The size of the expert teacher model is determined as follows. We first significantly increase the model size as much as possible (e.g., by at least 5×) and conduct a brief training phase to observe its performance. The implementation varies for different backbone models.

Specifically, a validation set is automatically generated in the original code of ALECE and FACE. We train these two backbones for 15 epochs (a full training takes 150 epochs) and observe their loss in the validation set. If the loss on the validation set at the checkpoint is close to or smaller than that of the small teacher model and rapidly decreases in the 15 epochs, we will keep the expert teacher model. As mentioned in our manuscript, Page 8, Experimental Setup- Number of teachers, sometimes "large models risk overfitting, leading to performance degradation". In that case, we will observe that the loss starts at a low point but decreases very slowly. We will adjust the model size (at most three different sizes) and track the performance until we observe the smallest loss at the checkpoint.

UAE and NeuroCard do not generate a validation set. Instead, these backbone models calculate the Q-error on the testing set at the end of each epoch. Thus, we train UAE and NeuroCard for three epochs (a full training takes 10 epochs) and monitor the 50th Q-error. If the 50th Q-error is smaller than that of the small teacher model and is still rapidly decreasing, we will keep the expert teacher model. If not, we will adjust the model size thrice and use the expert model with the smallest 50th Q-error.

1.2 Loss Difference Threshold ξ

We determined the loss difference threshold through an analysis of the training curves. We observed that the minimum loss of UAE, NeuroCard, and FACE is around 0.1, while ALECE is around 0.15. Based on these observations, we set the threshold to 0.1 for UAE, NeuroCard, and FACE, and 0.15 for ALECE to ensure optimal sample selection during distillation.