A. Experiments with no DT pretraining

In the following set of experiments, we pretrain only the IDM component of ALPT and not the DT. We show the finetuning performance results for the Narrow set of Atari games in Figure 6. Note that the axis here is up to 100k steps as opposed to 1M for the figures in the main text.

B. Implementation Details

In Table 3 we give the implementation details of our IDM and DT transformer architectures.

The IDM model is the same as the DT model, except that it is non-causal. This is enforced by changing the attention mask to a matrix of all 1 values in the IDM.

Table 3: A summary of the transformer model parameters.

Parameter	**Value**
Layers	6
Hidden Size	512
Heads	8
Batch Size	256
Weight Decay	5×10^{-5}
Learning Rate	3×10^{-4}
Gradient Clipping	1.0
β_1, β_2	$0.9, 0.999$
Warm-up Steps	4000
Optimizer	LAMB

C. Experiments with Conservative Q-Learning (CQL)

In this set of experiments, we examine the performance of Conservative Q-Learning (CQL) (Kumar et al., 2020) trained on a dataset of $10,000$ frames, as opposed to $500,000$ in the original work (Table 3 of CQL, 1% dataset size), from various Atari games utilized in our experiments. In Table 4 we report the final evaluation performance on the game after training for 100 iterations. All implementation details are consistent with the original implementation in the cited work. We utilize the CQL($\mathcal{H}$) method.

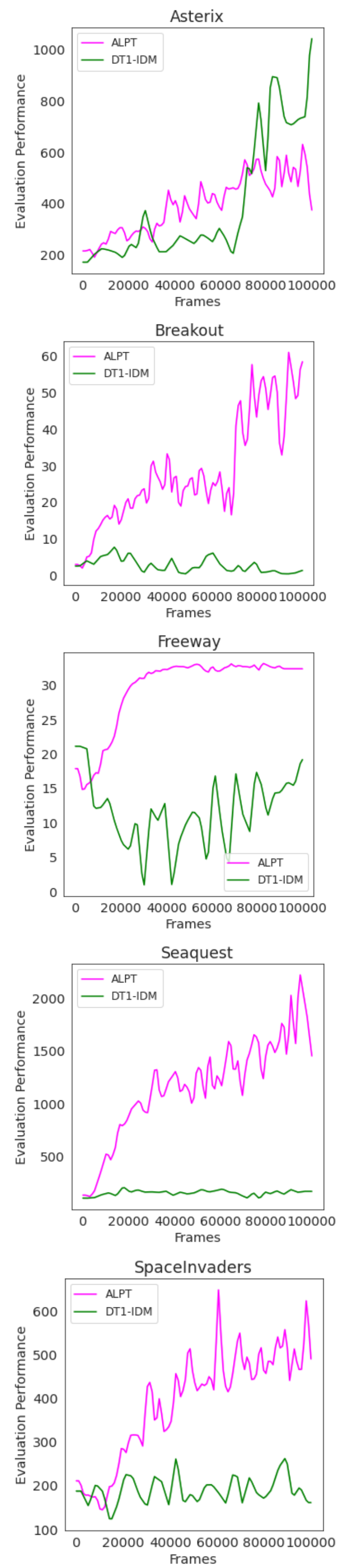

Figure 6: Evaluation game performance during finetuning of ALPT and DT1-IDM. In these experiments we do not pretrain the DT. $100k$ steps are shown.

Table 4: The final evaluation game performance after training CQL for 100 iterations on a dataset of 10,000 labelled frames from each Atari game.

Game Name	Final Performance
Asterix	227.5
Breakout	12.3
Freeway	10.2
Seaquest	236.0
SpaceInvaders	250.9

D. Source Code

We make the source code publicly available for our Maze experiment only at this time. The details can be found at: https://anonymous.4open.science/r/alpt_maze-5927/README.md.