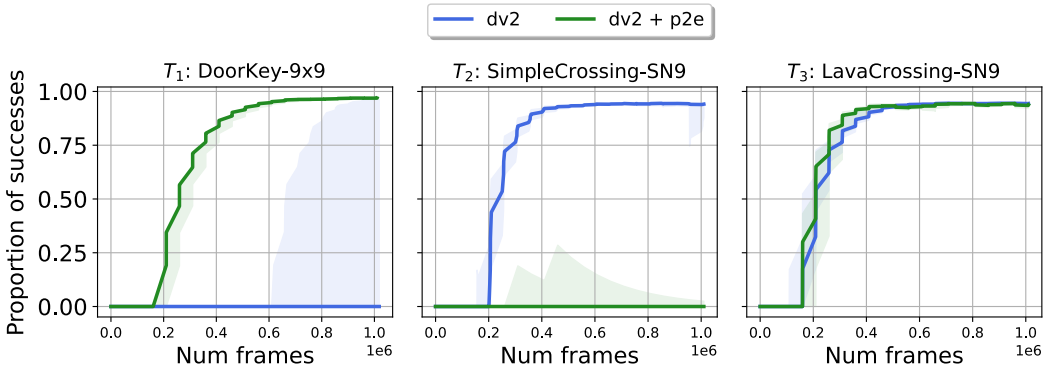

Figure 4: Single task performance of on individual tasks from the Minigrid CRL benchmark. All curves are a median and inter-quartile range over 5 seeds.

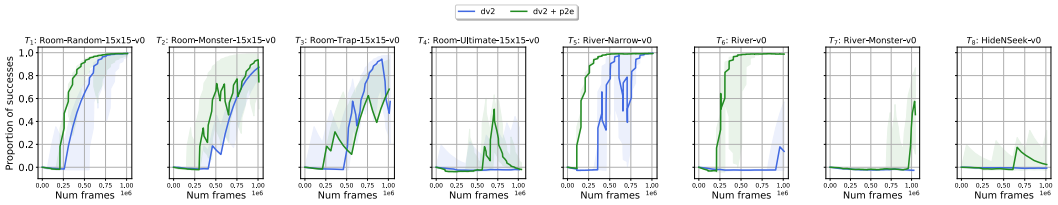

Figure 5: Single task performance of on individual tasks from the Minihack CRL benchmark. All curves are a median and inter-quartile range over 5 seeds.

Supplementary Material

Appendix A Single Task experiments

To assess the forward transfer of DreamerV2 for CRL we need the performance of each task as a reference Eq. (3). Single task learning curves for Minigrid are shown in Fig. 4 and single task learning curves for all Minihack tasks in the CRL loop are shown in Fig. 5.

Appendix B Further Experiments

A couple further experiments are introduced which are referenced in the main paper. In Appendix B.1 we explore various design choices required for DreamerV2 + Plan2Explore to get the best performance for CRL. Secondly, in Appendix B.2 we explore how increasing the size of the experience replay buffer size affects performance in the Minihack CRL benchmark.

B.1 DreamerV2 Ablation Experiments

We explore various design choices which come from the implementations of DreamerV2 [9] and Plan2Explore [13].

1. The use of Plan2Explore as an intrinsic reward.
2. World model learning by reconstructing the observations $\hat{o}_t$ only and not the observations, rewards and discounts all together.
3. The use of the exploration policy at to evaluate the performance on all current and past tasks rather than having a separate exploration and evaluation policy.

Plan2Explore	$\hat{o}$ reconstruction only	$\pi_{exp} = \pi_{eval}$	Avg. Performance (↑)	Avg. Forgetting (↓)	Avg. Forward Transfer (↑)
-	-	-	0.09 ± 0.07	0.37 ± 0.07	0.56 ± 0.86
✔	-	-	0.28 ± 0.13	0.13 ± 0.08	0.11 ± 0.15
✔	✔	-	0.39 ± 0.13	0.19 ± 0.16	0.87 ± 0.95
✔	✔	✔	0.38 ± 0.03	0.22 ± 0.05	0.76 ± 0.25

Table 3: CRL metrics for different design decisions on DreamerV2 for the Minihack CRL benchmark of 8 tasks. All metrics are an average and standard error over 5 seeds. ↑ indicates better performance with higher numbers, and ↓ the opposite.

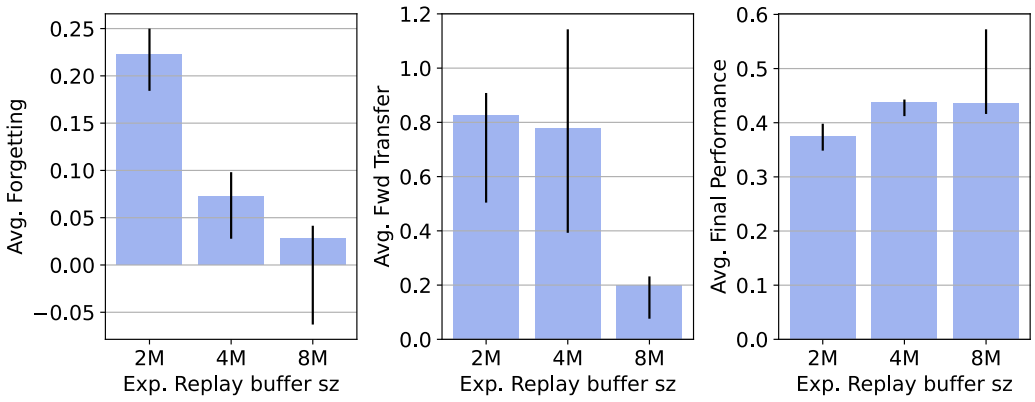

Figure 6: CRL metrics for DreamerV2 + Plan2Explore for the Minihack benchmark of 8 tasks versus the experience replay buffer size of the world model for DreamerV2 + Plan2Explore. All metrics are median and inter-quartile range over 5 seeds.

The results are shown in Table 3. We decided to pick the model in the final line to report the results in the main paper as they produce the good results on Minihack with relatively small standard errors.

B.2 Stability versus Plasticity: Increasing the Size of the Replay Buffer

By increasing the replay buffer size for world model learning for DreamerV2 + Plan2Explore we see that forgetting and average performance increases, however the forward transfer simultaneously decreases, Fig. 6. This is an instance of the stability-plasticity trade-off [57] in continual learning neural network based systems.