APPENDIX

The following Table I shows the hyper-parameters of each competence-based algorithm used in this research.

TABLE I: Per algorithm sets of hyper-parameters.

SMM hyper-parameters	**Value**
Skill dim.	4
Skill discriminator lr	10^{-3}
VAE lr	10^{-2}
DIAYN hyper-param.	**Value**
Skill dim	16
Skill sampling frequency	50 steps
Discriminator net arch.	512→1024→1024 →16 ReLU MLP
APS hyper-parameters	**Value**
Representation dim.	512
Reward transformation	$\log(r+1.0)$
Successor feat. dim.	10
Successor feat. net arch.	$\lvert\mathcal{O}\rvert$→1024→1024 →10 ReLU MLP
k in NN	12
Avg top k in NN	True
Least square batch size	4096

The Homeostasis Algorithm 1, which is defined in [1], is used for the evaluation implementation of .our model.

Algorithm 1 Homeostasis

Require:
  Target rate ρ
Initialize:
  $\overline{x} \leftarrow 0$, $\overline{x^2} \leftarrow 1$, , $\overline{x^+} \leftarrow 1$
for $t \in \{1, \ldots, T\}$ **do**
  obtain next scalar signal return x_t
  set time-scale $\tau \leftarrow \min(t, \frac{100}{\rho})$
  update moving average $\overline{x} \leftarrow (1-\frac{1}{\tau})\overline{x} + \frac{1}{\tau}x_t$
  update moving variance $\overline{x^2} \leftarrow (1-\frac{1}{\tau})\overline{x^2} + \frac{1}{\tau}(x_t - \overline{x})^2$
  standardise and exponentiate $x^+ \leftarrow \exp\left(\frac{x_t - \overline{x}}{\sqrt{\overline{x^2}}}\right)$
  update transformed moving average
    $\overline{x^+} \leftarrow (1-\frac{1}{\tau})\overline{x^+} + \frac{1}{\tau}x^+$
  sample $y_t \sim \text{Bernoulli}\left(\min\left(1, \rho\frac{x^+}{\overline{x^+}}\right)\right)$
end for

REFERENCES

[1] M. Pislar, D. Szepesvari, G. Ostrovski, D. Borsa, and T. Schaul, "When should agents explore?" *arXiv preprint arXiv:2108.11811*, 2021.