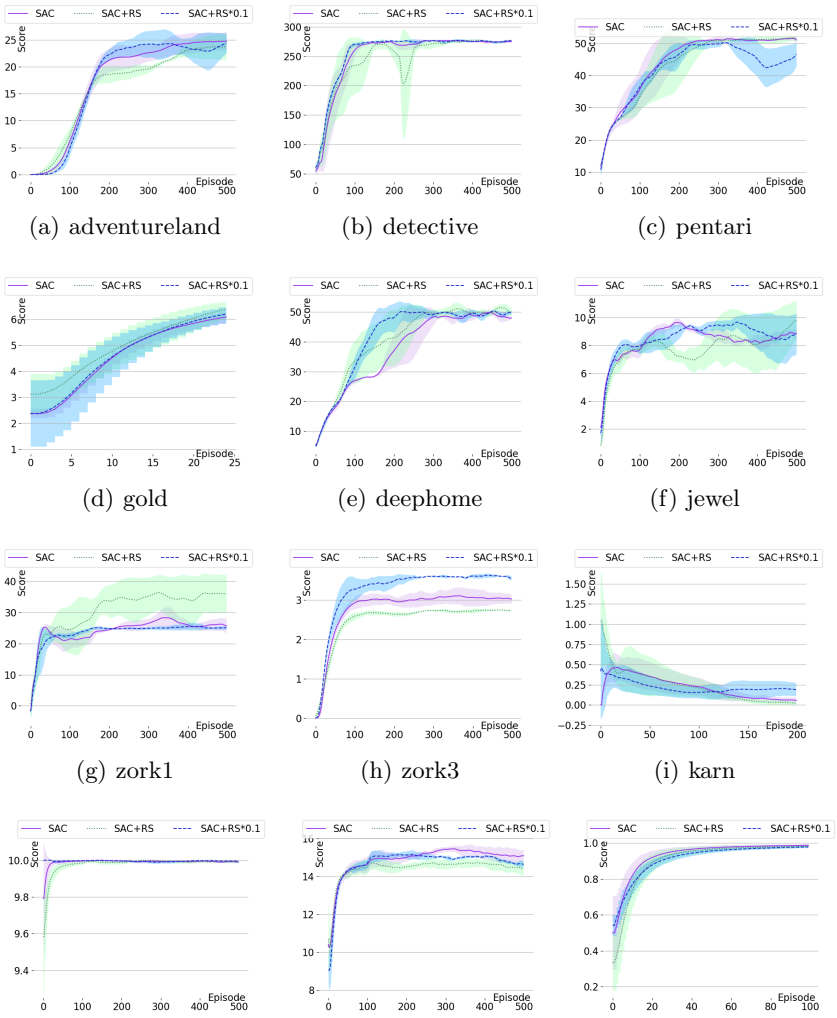
Supplementary Material: Learning to play text-based adventure games with maximum entropy reinforcement learning

Weichen Li[1], Rati Devidze[2], and Sophie Fellenz[1]

University of Kaiserslautern-Landau, Kaiserslautern, Germany
{weichen,fellenz}@cs.uni-kl.de
Max Planck Institute for Software Systems (MPI-SWS), Saarbrucken, Germany
rdevidze@mpi-sws.org

A Additional results

In Figure 1 we compare the update with SAC, SAC with Reward shaping (SAC+RS), SAC with re-scaled reshaped reward (SAC+RS*0.1). As the figure shows, convergence is faster with the reward-shaping technique for most games.



(a) adventureland (b) detective (c) pentari

(d) gold (e) deephome (f) jewel

(g) zork1 (h) zork3 (i) karn

(j) balances (k) ludicorp (l) yomomma

Fig. 1. This figure shows the development of the game scores over training episodes where shaded areas correspond to standard deviations. Compared is the SAC agent with and without reward shaping. You can see that reward shaping leads to faster convergence at the beginning of training for b) detective, e) deephome, d) gold, and h) zork3. The end score is higher with reward shaping for seven of twelve games compared to SAC agent. Reward shaping techniques are especially beneficial for difficult games compared to possible games.