1. Things seem to start working for both with and without negative rewards after removing the relative-position vector from the input state. (folder – 0)
2. Trained with positive rewards only (rewards clipped from 0 to 2), and reward for discovering new patch added on top of the clipped reward. This seems to learn to collect berries, and does collect about 156 berries in eval after episode 160. But after this the policy seems to degrade (learning rate too high? Or side effect of no rewards or side effect of positive rewards only?).