PPO + invalid action masking
+ selfplay + encoder-decoder