

A2C Hex Agent

Eberdorfer, Rodinger, Zlatanowski

Key Features

Pure Selfplay

2 Layer CNN -> Layer 1: 16 3x3 filters + ReLU, Layer 2: 32 3x3 filters + ReLU

Actor decides which move to make

Critic estimates chance of winning

Biggest training improvement -> data augmentation

For every agent move we 8x the data by rotating and mirroring the board

Agent is forced to learn the patterns of hex not just memorise positions

200k games simulated in selfplay -> average game around 20 moves * 200 000 * 8 (augmentation) = 32 million unique board positions

Only winning was rewarded no bridges, starting position, game length, ... to force the agent to develop a deep understanding of the game

Final agent is the winner of an “internal tournament” of our best checkpoints