Supervised and Reinforcement Learning from Observations in Reconnaissance Blind Chess

TIMO BERTRAM – JOHANNES KEPLER UNIVERSITY LINZ

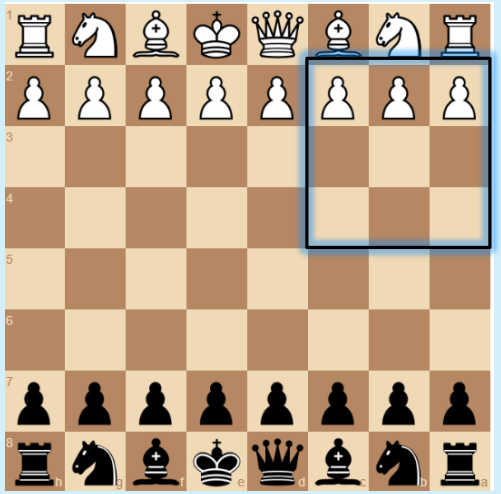
JOHANNES FÜRNKRANZ – JOHANNES KEPLER UNIVERSITY LINZ

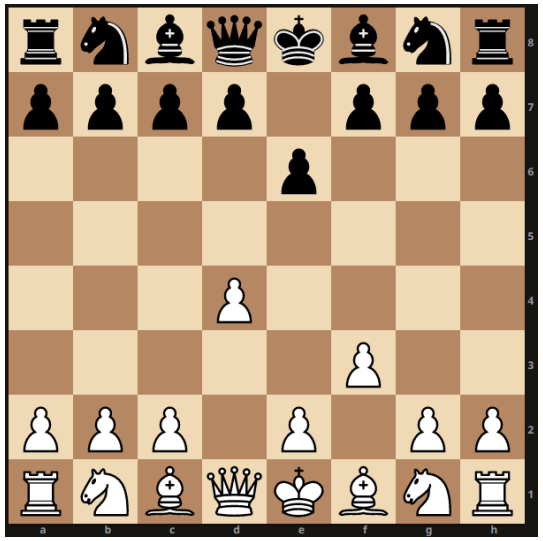
MARTIN MÜLLER – UNIVERSITY OF ALBERTA

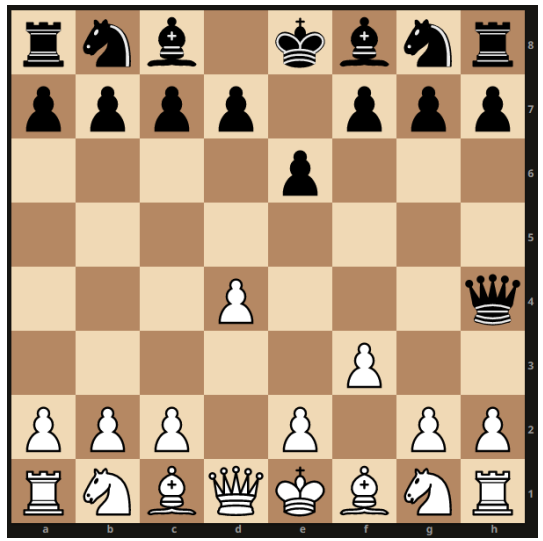
Board positions and logo obtained from https://rbc.jhuapl.edu/

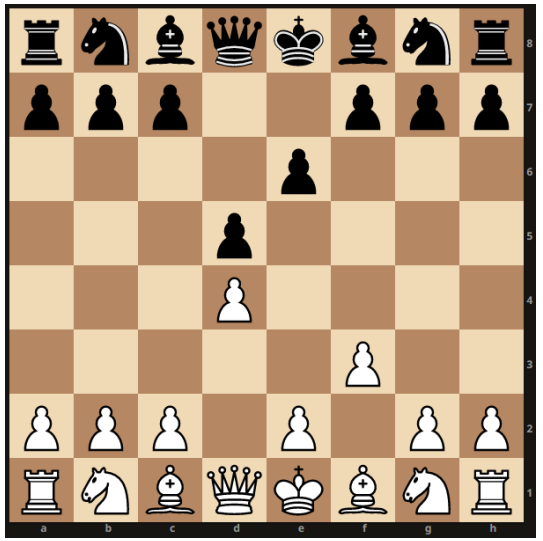
Reconnaissance Blind Chess

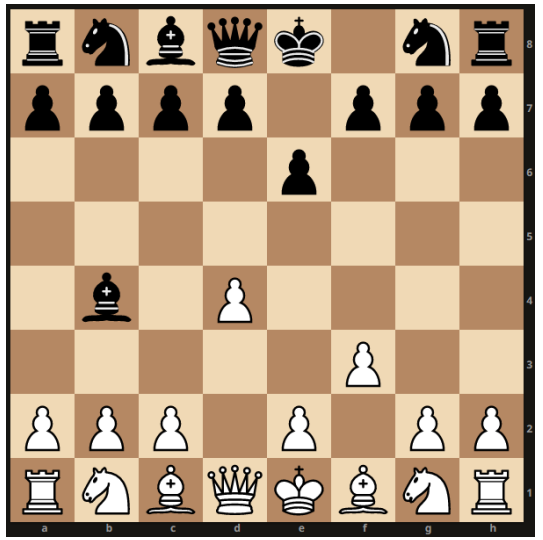
- Starts as a normal game of chess but information about opponent's moves is not obtained
- The majority of information is obtained by *sensing* a 3x3 area of the board each turn
- RBC significantly differs from chess
  - Legality of moves is generally unclear
  - *Checks* do not exist, the goal of the game is to capture the opponent's king

	Sense	Result	Move
1			
2			





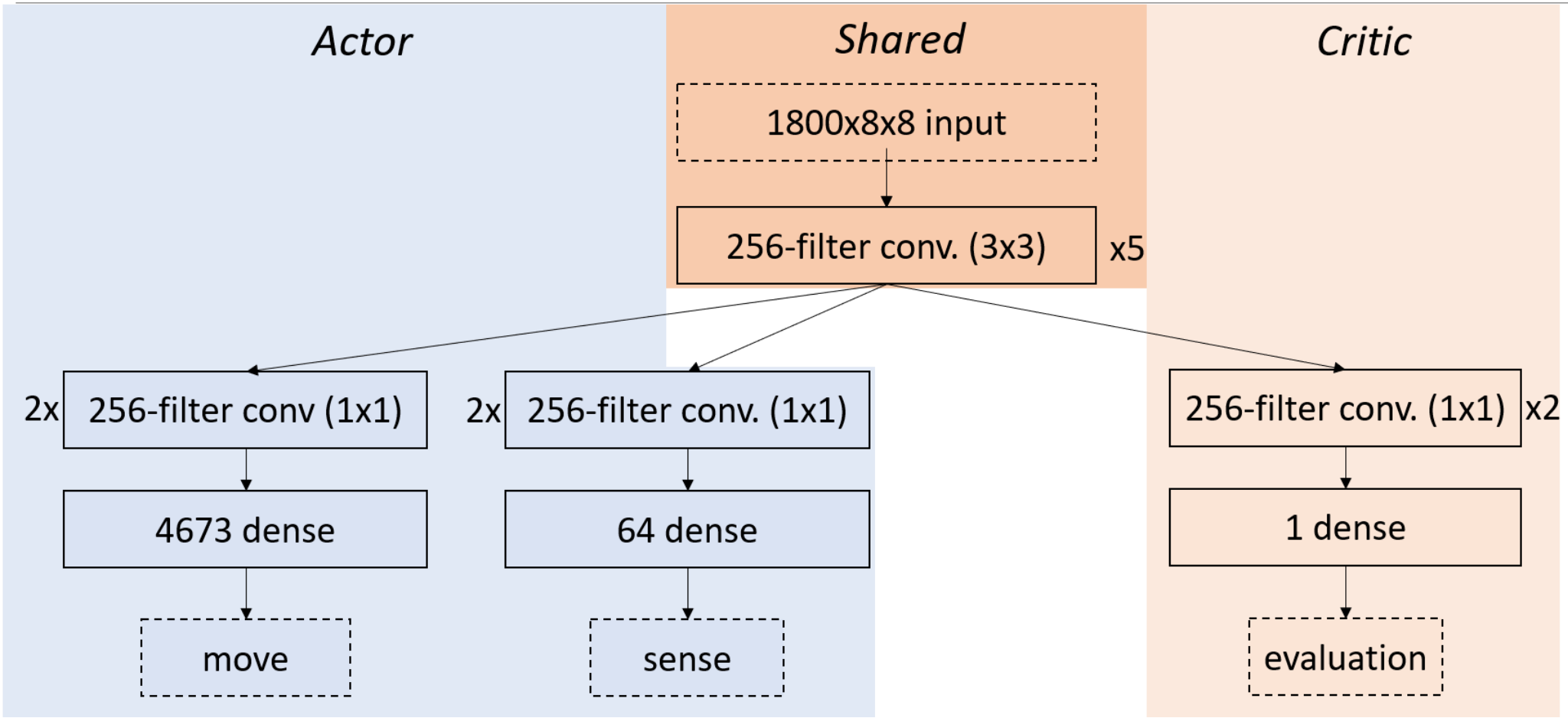




Our approach

- Previous work strongly focuses on reconstructing the true game state and then using a chess engine (Stockfish, LeelaZero) to make moves
  - It is not clear whether this is generally a good idea, as *good* moves in chess and RBC can be very different
- We aim to play this game in a different way and ask: *Do we actually need to know the game state or can we play directly from the obtained observations?*
- We train a neural network to parameterize a policy based on a history of observations

Network structure



RECONNAISSANCE BLIND CHESS

Training

Phase 1: Supervised Learning

- We use expert games and train the network to predict the taken actions (Actor) and the result of the game (Critic)
- This should result in learning some basic concepts of the game, but:
  - Lots of conflicting information (different actions taken in the same situation)
  - No special importance in capturing the opponent's king

Phase 2: Reinforcement Learning

- Let the agent play against itself to improve the quality of moves
- We use PPO to further train the pre-trained Actor and Critic networks
  - Agent is incentivized to actually win the game, as this is the only reward signal

Results

Phase 1: Supervised Learning

- The network learns to predict the actions from the database
- Predictive test-accuracy of 49%
- Elo of 1118

Phase 2: Reinforcement Learning

- Agent constantly learns to beat previous versions of itself
- Elo of 1330

Takeaways

Can we learn to play the game without reconstructing the full game state?

- **Yes! We trained a neural network to play the game only based on observations**
- **We see RBC-specific policies**
- **Not state-of-the-art but promising proof-of-concept (rank 27 out of 84)**

RECONNAISSANCE BLIND CHESS

Questions?

TBERTRAM@FAW.JKU.AT