NEURAL INFORMATION PROCESSING SYSTEMS

Game Solving with Online Fine-Tuning

Ti-Rong Wu,[1] Hung Guei,[1] Ting Han Wei,[2]
Chung-Chin Shih,[1,3] Jui-Te Chin,[3] I-Chen Wu[1,3]

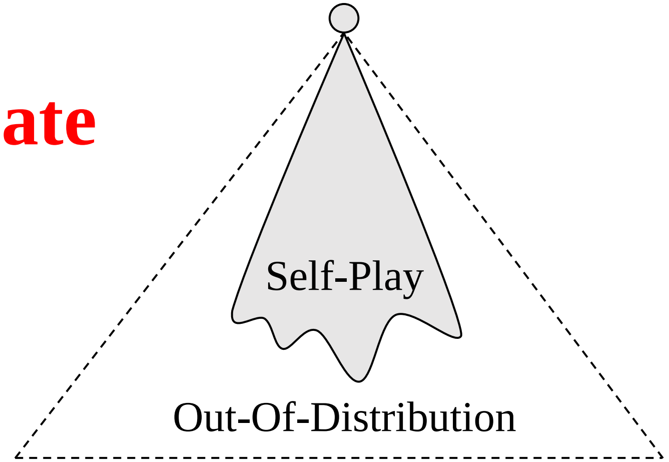
[1] Academia Sinica
[2] University of Alberta
[3] National Yang Ming Chiao Tung University

Game Solving with AlphaZero

- AlphaZero not only demonstrates super-human levels in game playing, but also **serves as heuristics in game solving**

- To solve a game, a winning response must be found for all possible moves by the losing player, which **includes very poor lines of play**

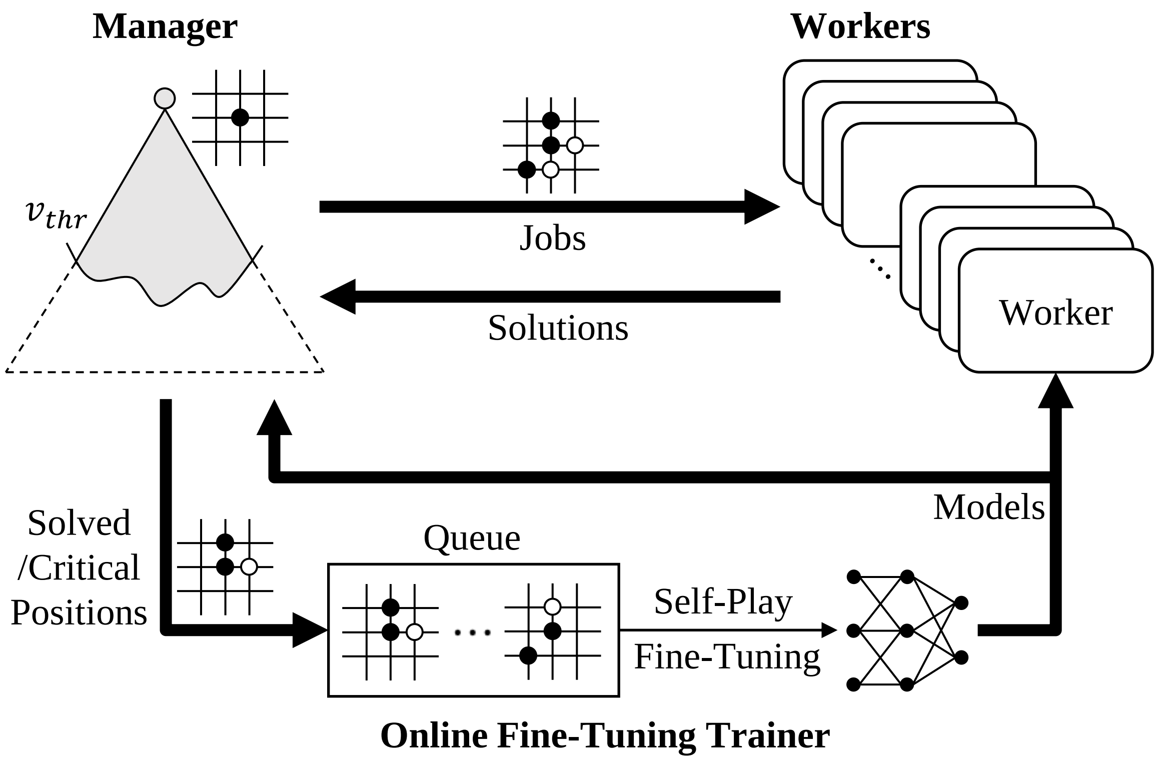
  ⟹ **For game solving, the fixed, pre-trained AlphaZero heuristics can be highly inaccurate**



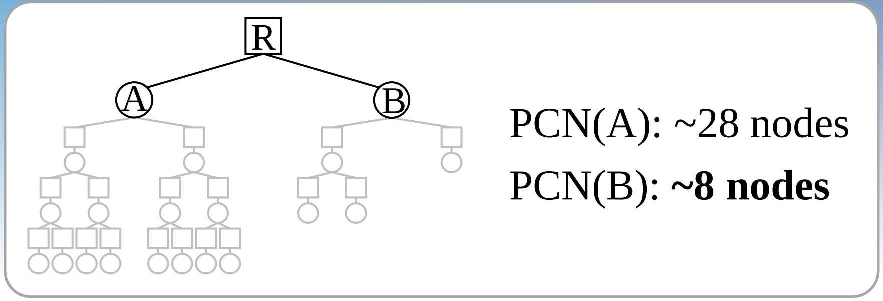
Game Solving with Online Fine-Tuning

- We investigate **online fine-tuning to** **learn tailor-designed heuristics**

- The **online fine-tuning solver** comprises three components:
  - *Manager*
  - *Workers*
  - *Online Fine-Tuning Trainer*

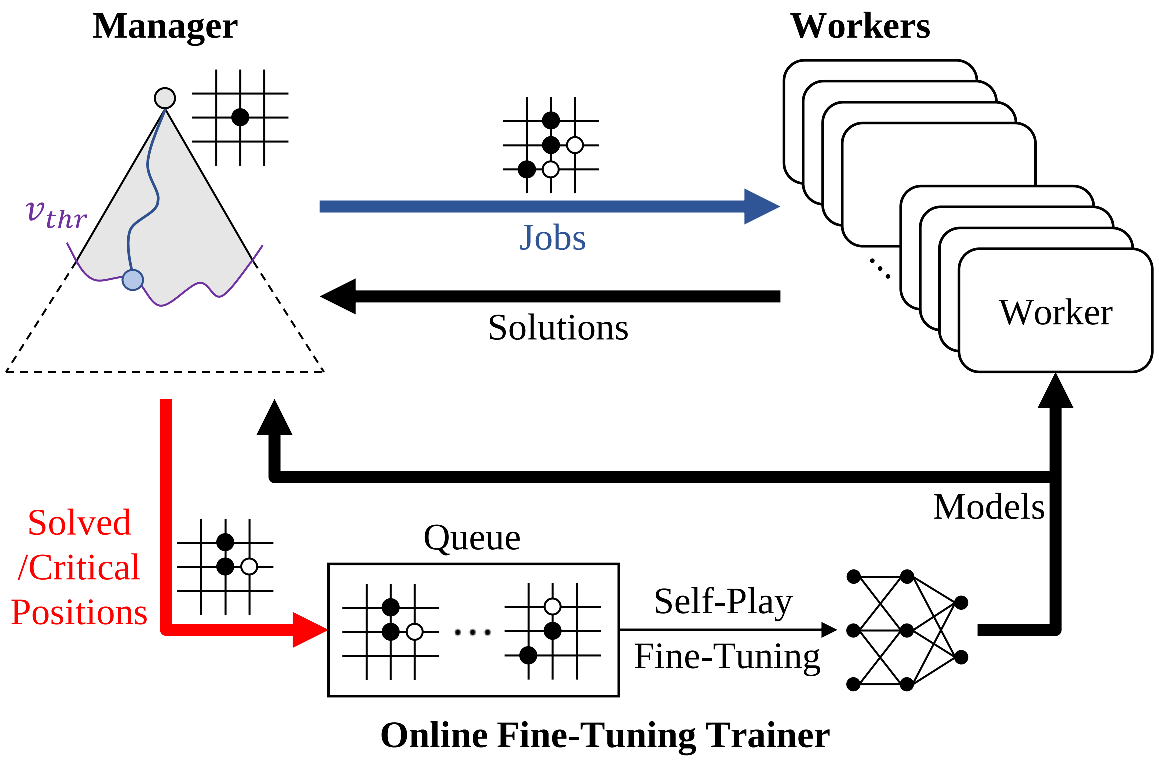


Manager



PCN(A): ~28 nodes
PCN(B): **~8 nodes**

- Maintain the proof search tree
  - Perform *Monte Carlo tree search*
- Employ a heuristic to **assign jobs to workers** to solve
  - Use *Proof Cost Network* to predict the cost for solving the position
  - Determine whether to assign to workers by a *cost threshold* v_{thr}
- Forward **solved/critical positions** to online fine-tuning trainer


Manager

Workers

v_{thr}

Jobs

Solutions

Worker

Models

Solved /Critical Positions

Queue

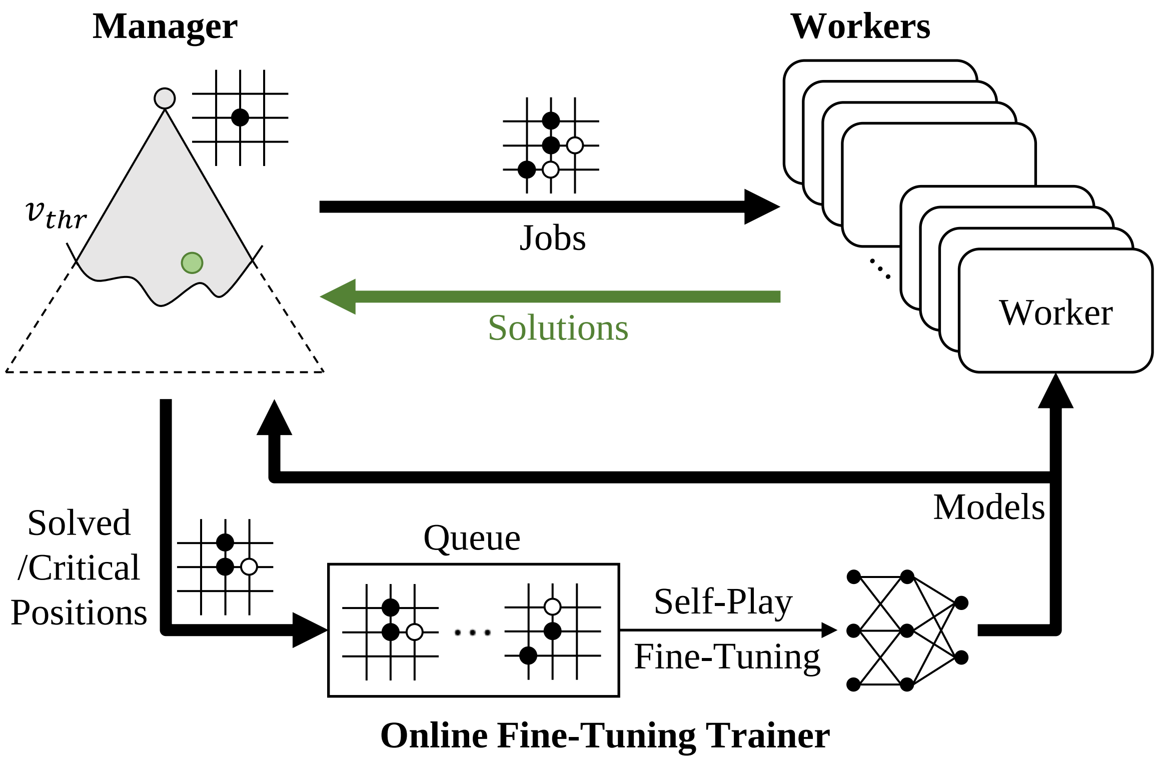
Self-Play Fine-Tuning

Online Fine-Tuning Trainer

Institute of Information Science, Academia Sinica
中央研究院 資訊科學研究所

Workers

- Attempt to solve the jobs in parallel
  - Within given computing constraints
  - Employ the same heuristic as manager
- **Return the solutions to manager**

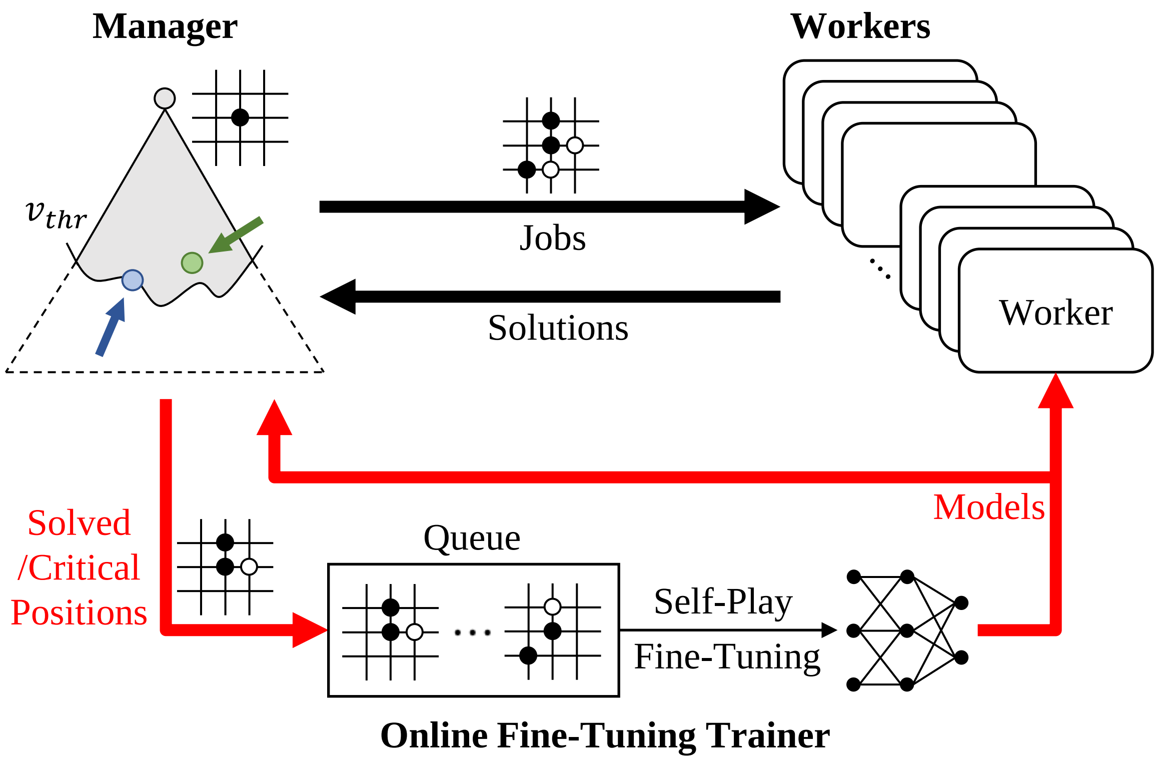


Online Fine-Tuning Trainer

- **Fine-tune the heuristic** using
  - **Solved positions**: where theoretic outcomes are found

    ⟹ Guide the model to learn their theoretic outcomes
  - **Critical positions**: where manager is currently focused on

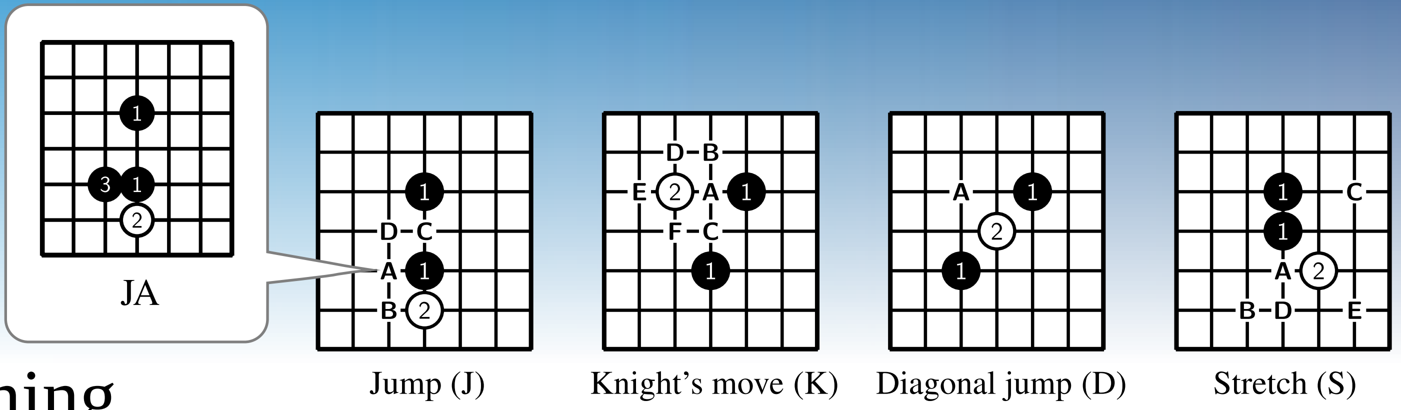
    ⟹ Use them as initial positions to perform self-play
- **Update the heuristic** employed by manager and workers



Proof Cost of ● = 0

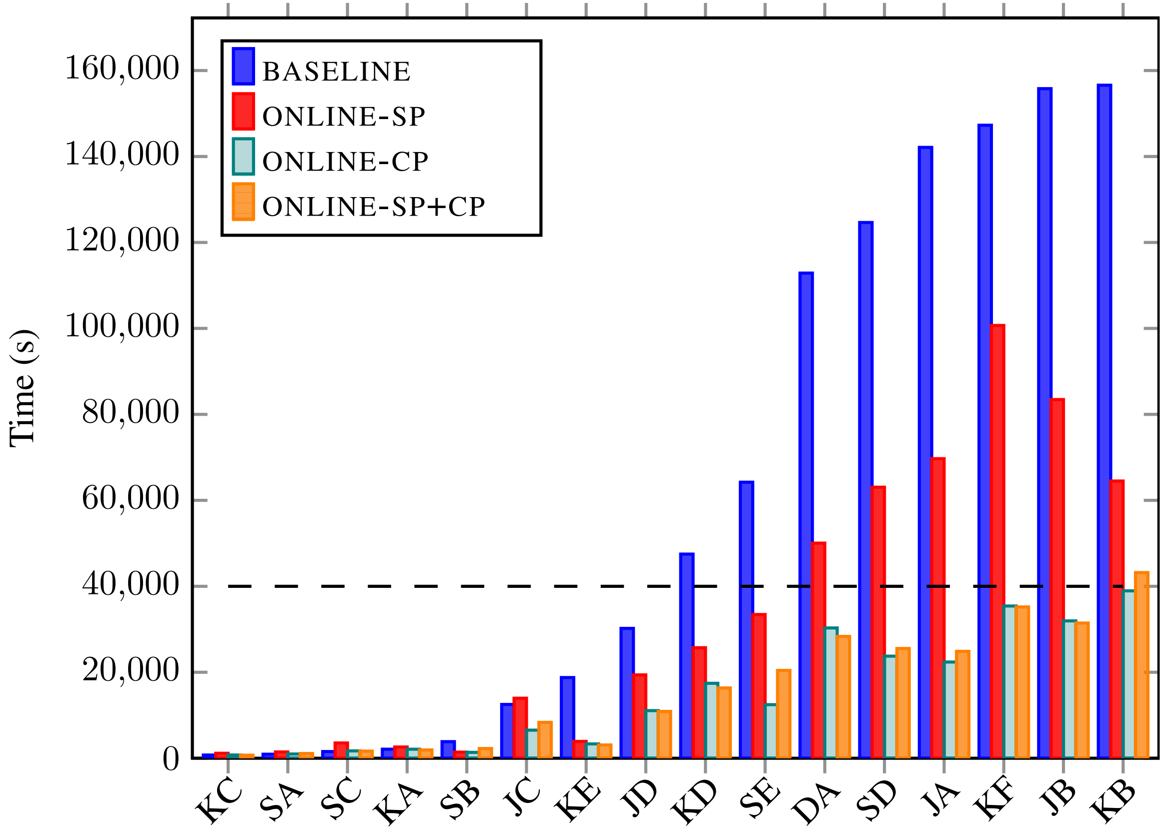
Self-Play Begins from ●

Experiments



- We evaluate online fine-tuning solvers on 16 challenging 7x7 Killall-Go opening problems
  - SP: w/ solved positions
  - CP: w/ critical positions

- In general, online fine-tuning solvers significantly reduce the solving time by **using only 23.54% of computation time**



Summary

- Pre-trained AlphaZero-based models provide less accurate heuristics

  ⟹ Not optimal for solving problems

- **Online fine-tuning solvers learn tailor-designed heuristics**
  - Dynamically during solving
  - According to the manager’s attention

  ⟹ Find faster solutions

- We expect it has the potential to extend to
  - Single-player games such as Rubik’s Cube
  - Even other non-game fields

Thank You for Your Attention

Our code and data are available at
https://rlg.iis.sinica.edu.tw/papers/neurips2023-online-fine-tuning-solver