



# Transferable Reinforcement Learning for Board Games

*Natthaphon Hongcharoen*

Thesis presented for Faculty of Engineering and Technology  
Panyapiwat Insitute of Management  
In study of bachelor of science, Computer Engineering  
Study year 2018

# Abstract

The modern Reinforcement Learning has become widely interested recently, after the AlphaGo[4] became the first program to defeat a world champion in the game of Go. And by 2017, the AlphaGoZero[6] and AlphaZero[5] programs achieved superhuman performance in the game of Go and Chess, by solely trained from games of self-play which require no human knowledge. But in contrast, both AlphaGoZero and AlphaZero required extremely powerful processor as they need to random move in the early state of training which cost more expend. While in Computer Vision domain, we often use pretrained model from large dataset such as Imagenet[1] and retrain it for desired task which cost less time and achieved more accuracy than train it from scratch. In this thesis, we experiment a method to reuse the trained model of a game such as Othello, Connect4 or Gomoku and re-training with one different game by hoping it to be faster.

# Chapter 1: Introduction

## 1.1 Background

The modern Reinforcement Learning such as AlphaGo has showed outstanding performance by won against world champion in game of Go a decade earlier than predicted[2][18]. And the AlphaGoZero has showed that it possible to train Reinforcement Learning without any human knowledge while still achieved good performance[6][5]. In contrast of good result, for AlphaGo, by using ‘supervised learning from human expert games’[4] means it's require the expert games that might be difficult to obtained or inthe worst case, it might be impossible to get. And in case of AlphaGoZero, according to original paper[6], the model need to be trained with 29 million games of self-play. Parameters were updated from 3.1 million mini-batches of 2,048 positions each. In Gian-Carlo Pascutto's experiments, every 2000 self-play moves generated by GTX-1080ti GPU would take 93 seconds, which means it need 1700 years to play all 29 million games[3]. It practically impossible to train AlphaGoZero in personal level.

Meanwhile in Computer Vision domain, the Transferring weights which pretrained in large datasets is widely used as there are ImageNet[1] pretrained weights in many popular libraries[15][16][17] because of it often improve regularizing and better performance and need less sample to train[14]. In the same way, Reinforcement Learning algorithm should be able to transferring features which makes training process faster.

## References

- [1] DENG, J., DONG, W., SOCHER, R., LI, L.-J., LI, K., AND FEI-FEI, L. Imagenet: A large-scale hierarchical image database. *2009 IEEE Conference on Computer Vision and Pattern Recognition* (2009).
- [2] LEVINOVITZ, A. The mystery of go, the ancient game that computers still can't win. <https://www.wired.com/2014/05/the-world-of-computer-go/>, 2014.
- [3] PASCUTTO, G.-C. [computer-go] zero performance. <http://computer-go.org/pipermail/computer-go/2017-October/010307.html>, 2017.
- [4] SILVER, D., HUANG, A., MADDISON, C. J., GUEZ, A., SIFRE, L., VAN DEN DRIESSCHE, G., SCHRITTWIESER, J., ANTONOGLOU, I., PANNEERSHELVAM, V., LANCTOT, M., DIELEMAN, S., GREWE, D., NHAM, J., KALCHBRENNER, N., SUTSKEVER, I., LILICRAP, T., LEACH, M., KAVUKCUOGLU, K., GRAEPEL, T., AND HASSABIS, D. Mastering the game of go with deep neural networks and tree search. *Nature* 529 (2016), 484–489.
- [5] SILVER, D., HUBERT, T., SCHRITTWIESER, J., ANTONOGLOU, I., LAI, M., GUEZ, A., LANCTOT, M., SIFRE, L., KUMARAN, D., GRAEPEL, T., LILICRAP, T. P., SIMONYAN, K., AND HASSABIS, D. Mastering chess and shogi by self-play with a general reinforcement learning algorithm. *CoRR abs/1712.01815* (2017).
- [6] SILVER, D., SCHRITTWIESER, J., SIMONYAN, K., ANTONOGLOU, I., GUEZ, A. H. A., HUBERT, T., BAKER, L., LAI, M., BOLTON, A., CHEN, Y., LILICRAP, T., HUI, F., SIFRE, L., VAN DEN DRIESSCHE, G., GRAEPEL, T., AND HASSABIS, D. Mastering the game of go without human knowledge. *Nature* 550 (2017), 354–359.