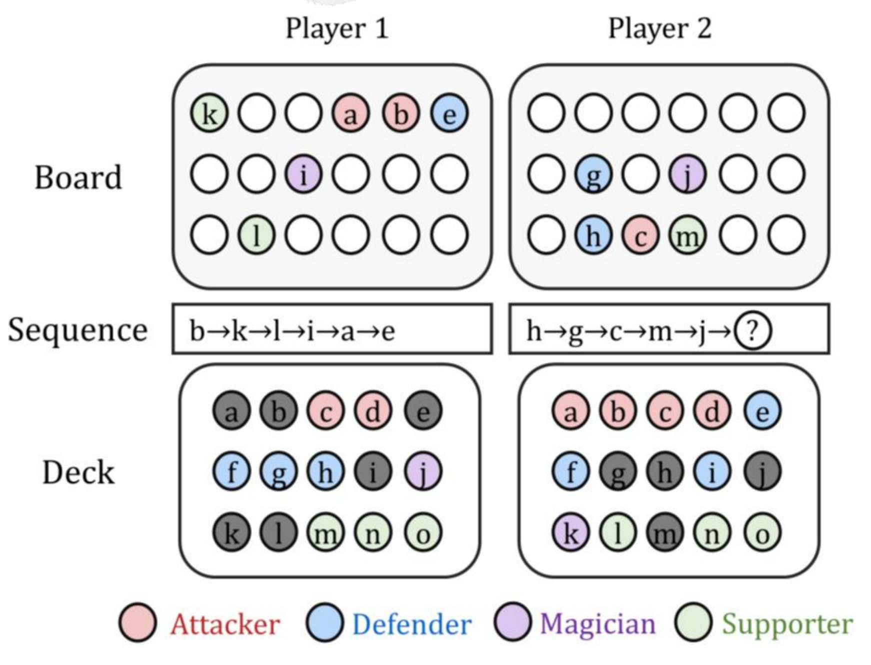
Conquering a rule-changing game with the action relevance aware AlphaZero

Hojoon Lee, Dongyoon Hwang, Jeesoo Woo and Jaegul Choo, Korea University

MOTIVATION

BrownDust



Game Rule

- 2-player turn-based game
- Each player has a deck of 15 characters.
- Players place 9 characters on each board one after another.
- After all 18 characters have been placed on the board, characters take action by the order they have been placed

Game Complexity

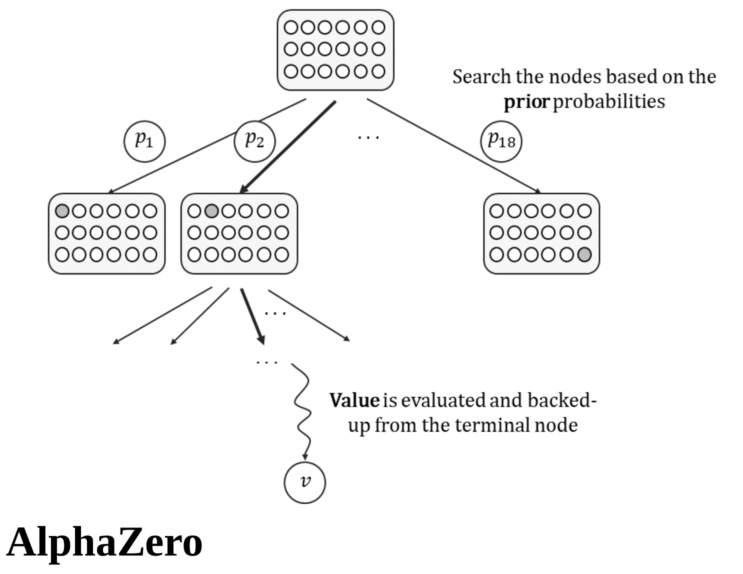
- Go(10^{170}) >**BD(10^{74})** >Chess (10^{47})

Problems

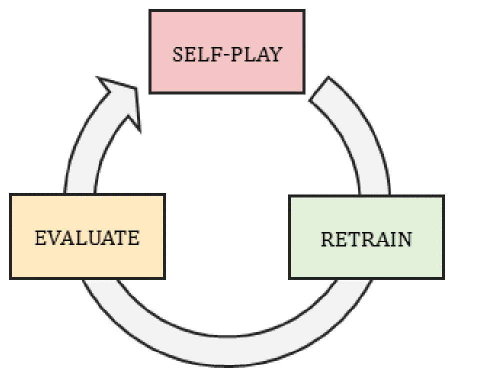
- New characters are updated regularly, requiring the model to be retrained
- Huge action space makes the training process infeasible

RELATED WORK

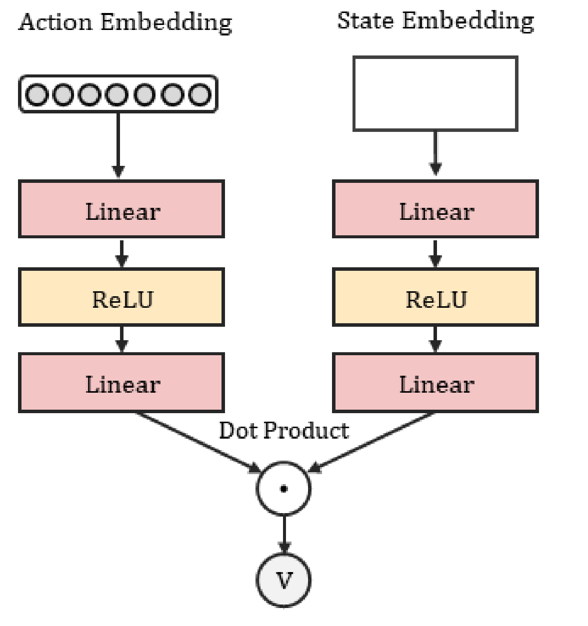
Monte Carlo Tree Search



AlphaZero



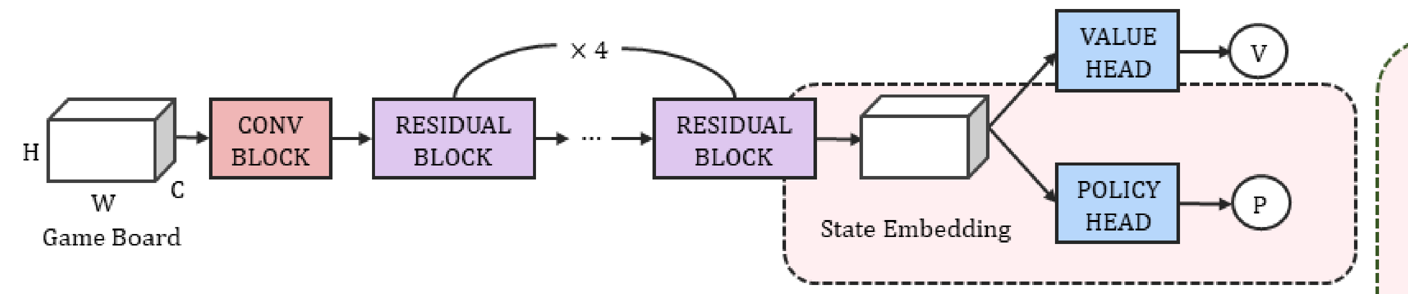
Deep Relevance Network

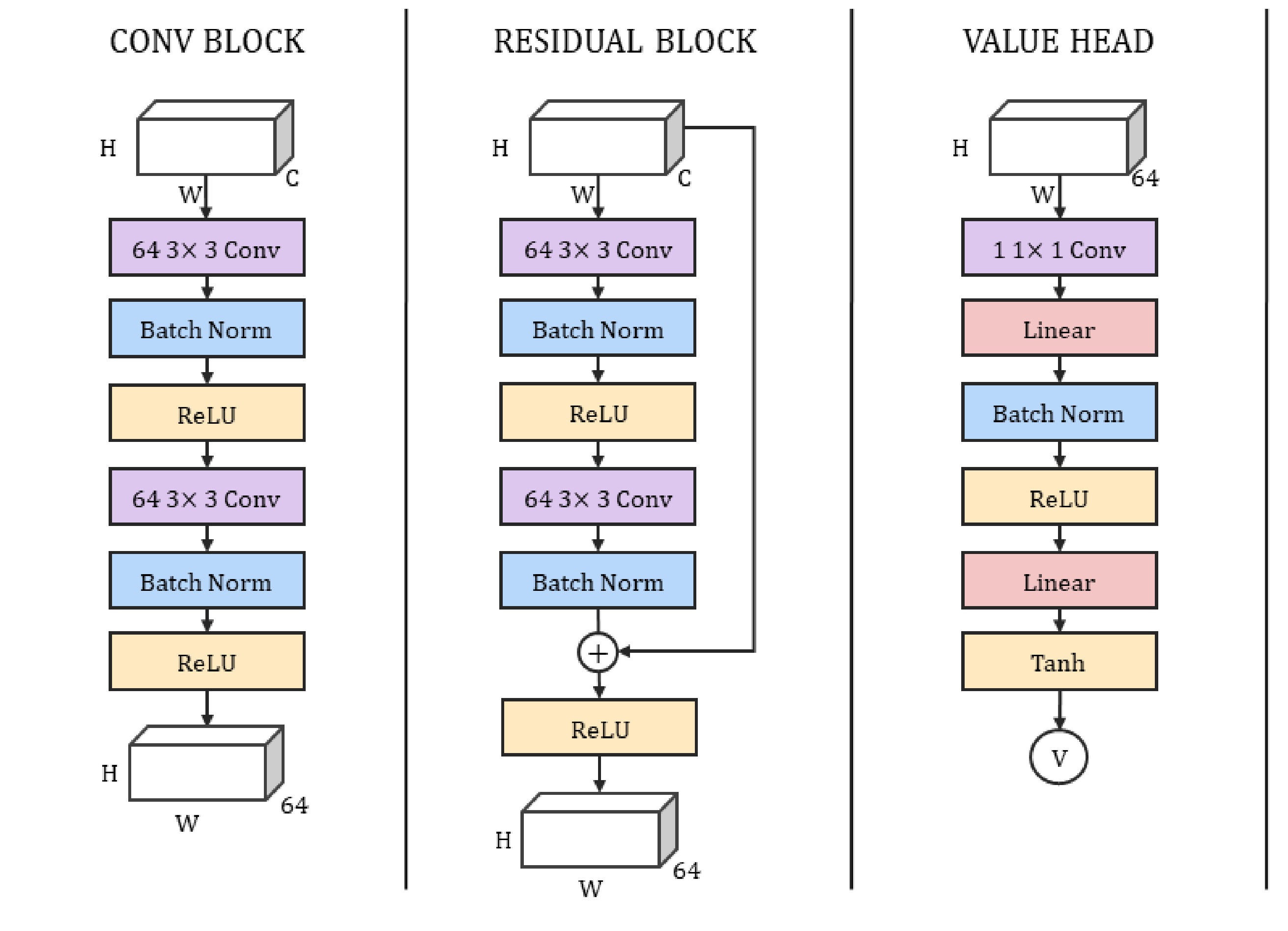


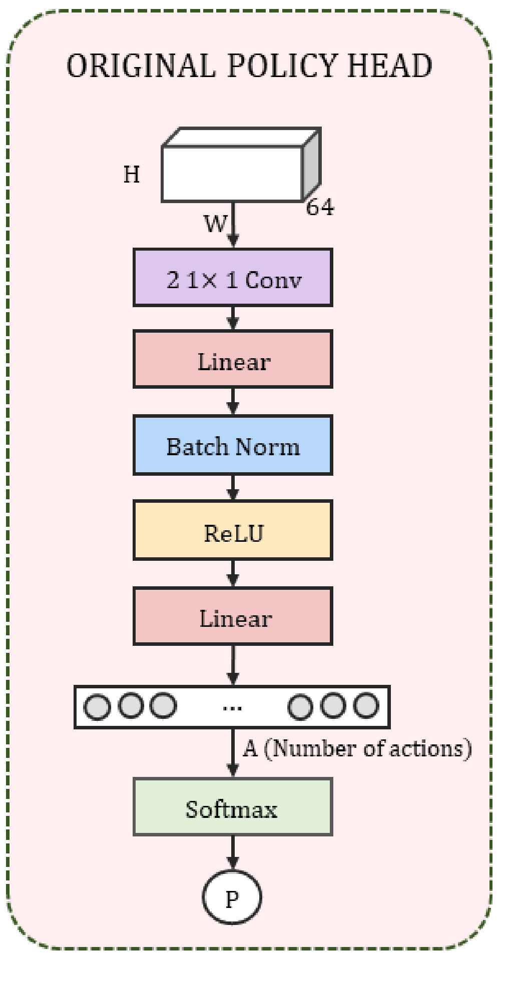
- **Self-Play** : Create a 'training set', by having the current best model play multiple games against itself
- **Retrain** : Optimize the neural network weights, by sampling a mini-batch from the self-play results
- **Evaluate** : Evaluate the retrained model by playing multiple games with the previous models

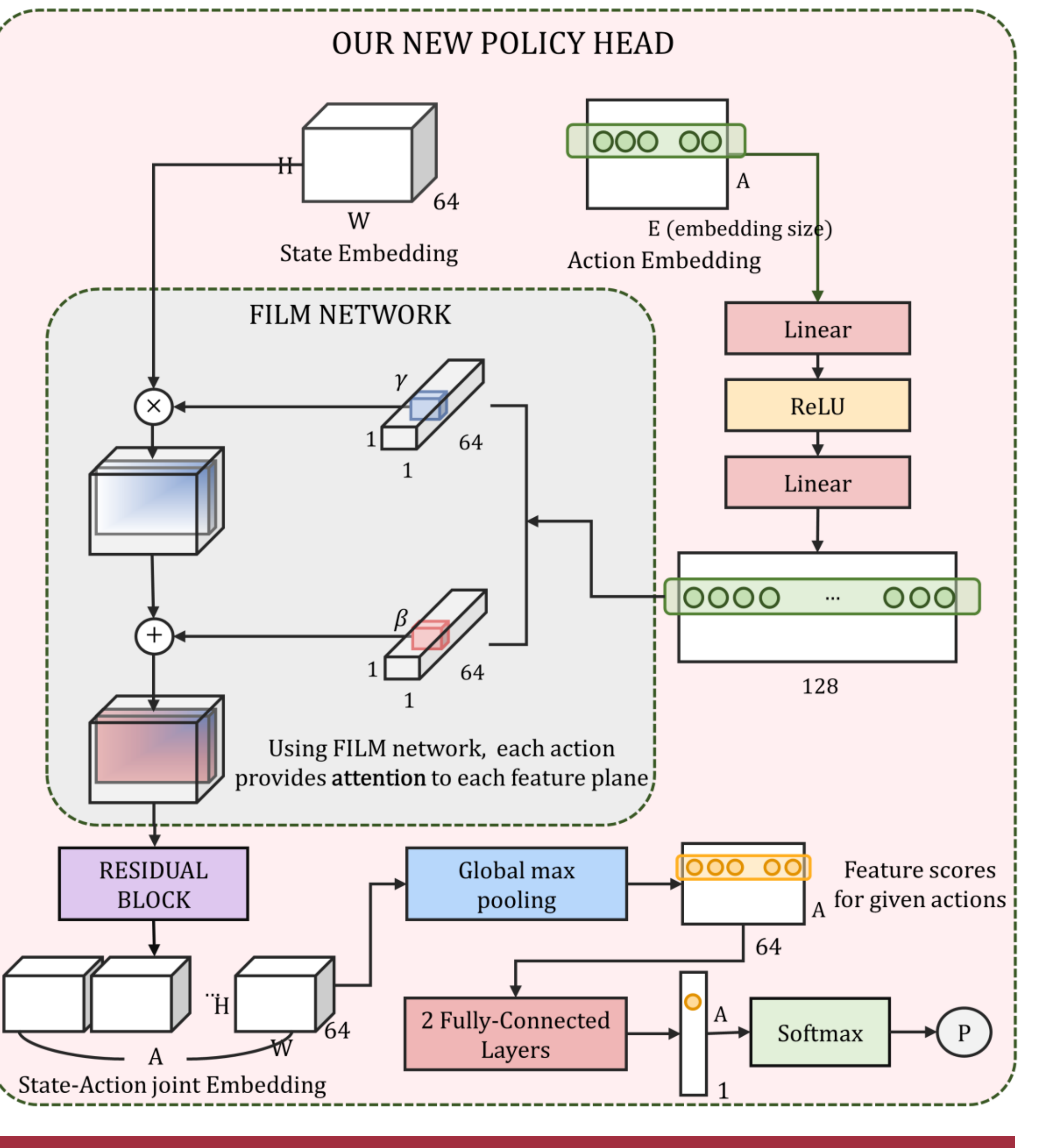
METHODS

Model Architecture



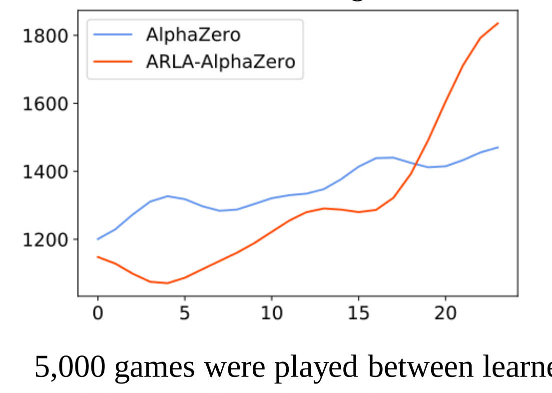




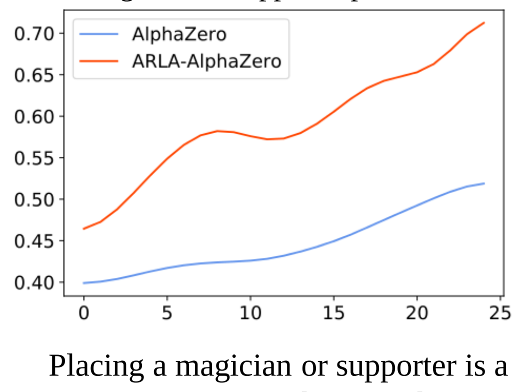


EVALUATION

Performance

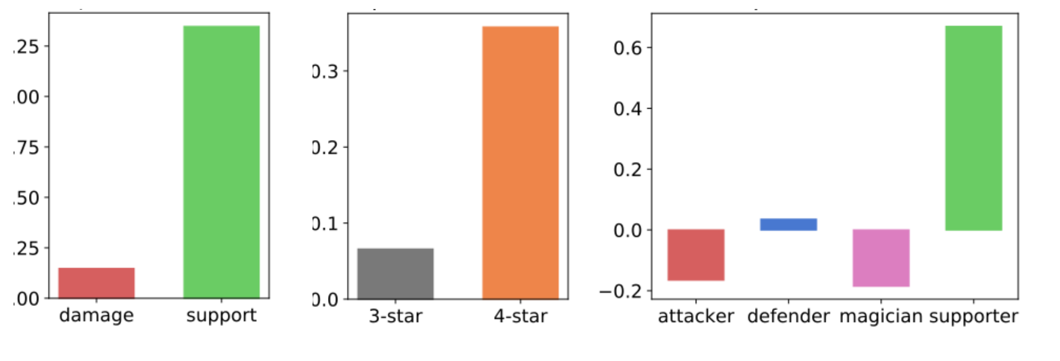


5,000 games were played between learned models. Based on the results, ELO ratings were estimated using the Bayesian method



Placing a magician or supporter is a prominent strategy between human experts in the early stages. ArlaZero quickly converges to this strategy

Analysis



By visualizing the action embedding features weights, we can confirm that the network learned emphasizing 4-star characters compared to 3-star characters further exploiting supporters than the other classes

Computational Complexity

	Space	Time
AlphaZero	O(AH)	O(AH)
ArlaZero	O(H)	O(SAH)

S : state size A: action size H : hidden size

- **Agreeable action space** : ArlaZero's complexity is dominated by the residual blocks
- **Huge action space :** Though Alphazero's excessive number of action parameters make training infeasible, ArlaZero's space complexity is robust to the action size by sharing parameters between actions

CONCLUSION

Summary

- By taking advantage of action relevance information, the learning process becomes much more efficient
- The model is robust to unseen actions with prior knowledge

Future Work

- Enhanced scalability of the FILM module will work nicely with bulky action spaces
- Due to lack of resources, our experiment of this model is the downsized version of the original AlphaZero. The results of our model given the complexity of the original is called for

REFERENCES

[1] Ji, H., Jianshu, C., Xiadong, H., Jianfeng, G., Lihong, L., Li, D., & Mari, O. Deep Reinforcement Learning with a Natural Language Action Space. In Association for Computational Linguistics (ACL) , 2016.

[2] Silver, D., Schrittwieser, J., Simonyan, K. *et al.* Mastering the game of Go without human knowledge. In *Nature* **550,** 354–359 , 2017

[3] Ethan, P., Florian, S., Harm, D., V., Vincent, D., & Aaron, C. FiLM: Visual Reasoning with a General Conditioning Layer. In Association for the Advancement of Artificial Intelligence(AAAI), 2018