AI for Competitive Play in Terra Mystical

Luis Antonio Perez

Computer Science Department, Stanford University
CS 221: Artificial Intelligence

Stanford University

Motivations and Goal

Problem and Goal
- Create an AI agent capable of playing TM at super-human level.

What is Terra Mystical (TM)
- 2-5 player strategic board game with collaborative components.
- Each player tries to maximize the number of victory points (VP) earned during six fix rounds of gameplay,

Why TM
- Recent breakthroughs in strategy games (Go and Chess/Shogi)
- Complex gameplay mechanics, multiple "actions" per game
- No current AI has beaten top human-players
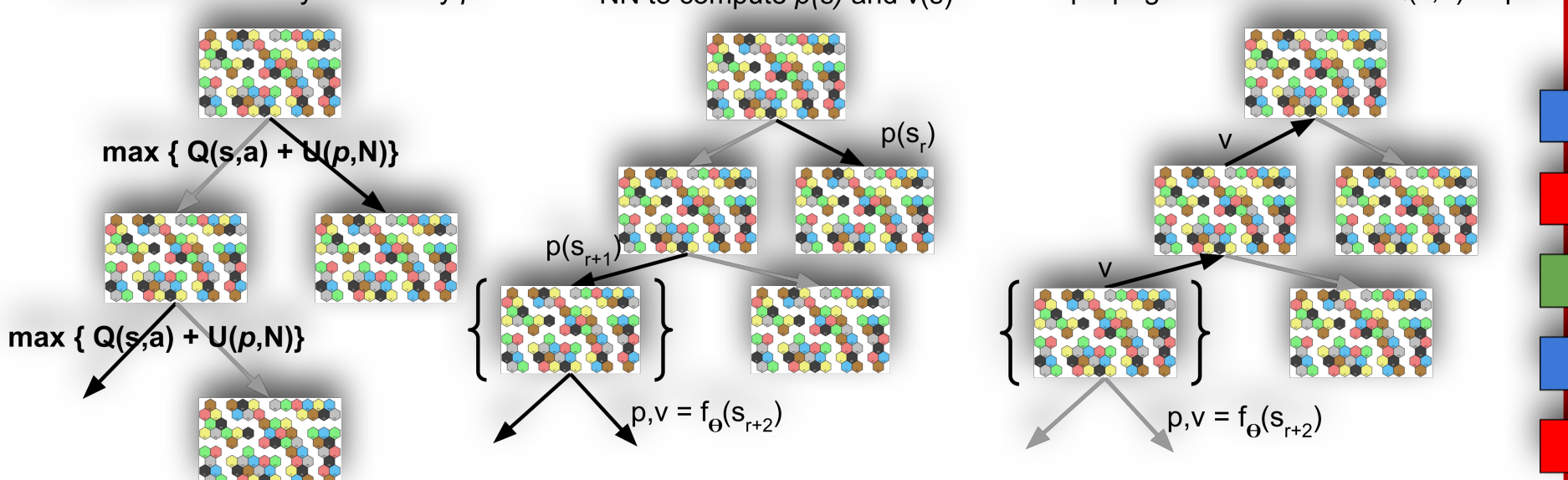

Monte Carlo Tree Search

SELECTION
Traverse Existing Search Tree with Deterministic Policy Guided by p

EXPANSION + EVALUATION
On non-terminal leaf-node, expand using NN to compute $p(s)$ and $v(s)$

BACKPROPAGATION
Use estimated $v(s)$ for leaf-state and propage this result to all $Q(s,a)$ of path

$\max \{ Q(s,a) + U(p,N)\}$

$p(s_r)$

$p(s_{r+1})$

$\max \{ Q(s,a) + U(p,N)\}$

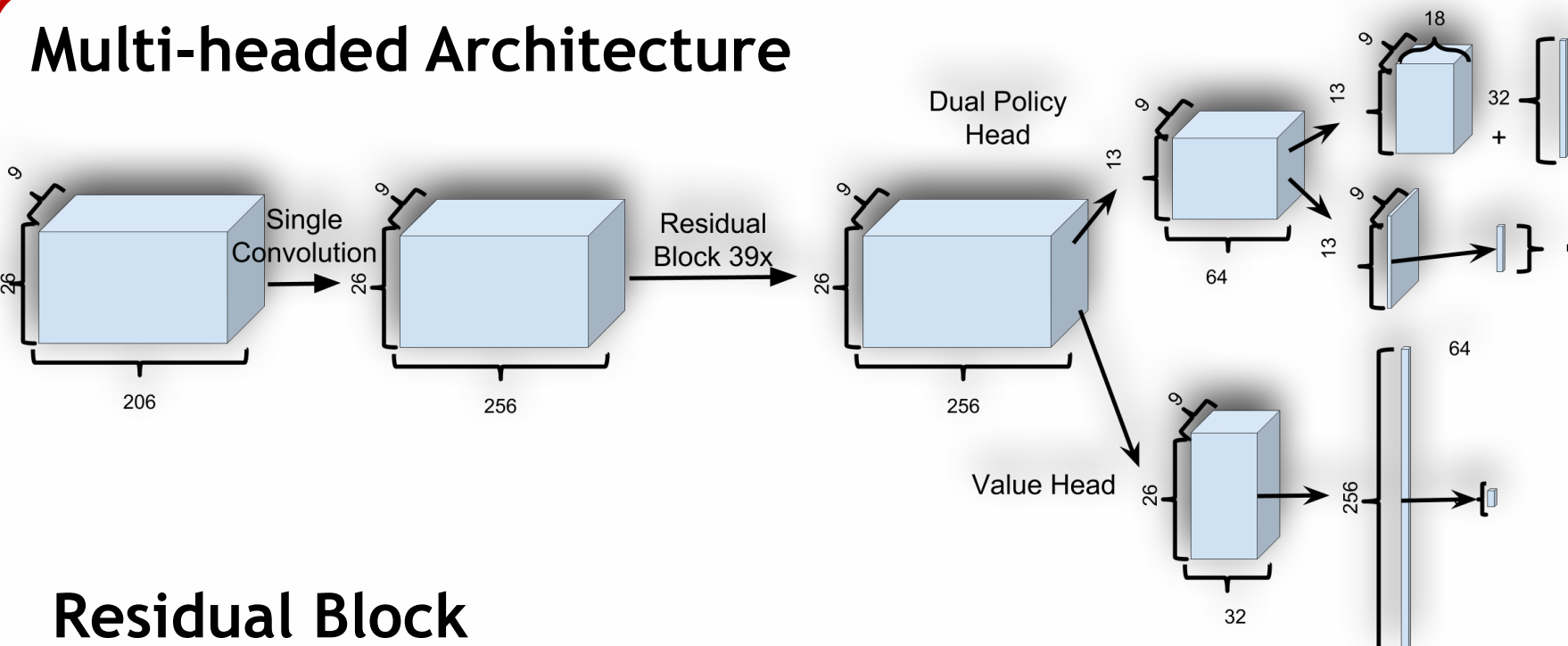$p,v = f_\theta(s_{r+2})$

$p,v = f_\theta(s_{r+2})$

Self-Play Reinforcement Learning

- **Value Estimation** - for newly encountered states, estimate value and use this as proxy for full MCTS rollout
- **Neural Network Guided MCTS search** - prior probabilities set by NN predictions
- Loss Function
  - Collect set of **pi**, z and use to update NN predictions
  - Can occur while performing self-play - parameters updated immediately

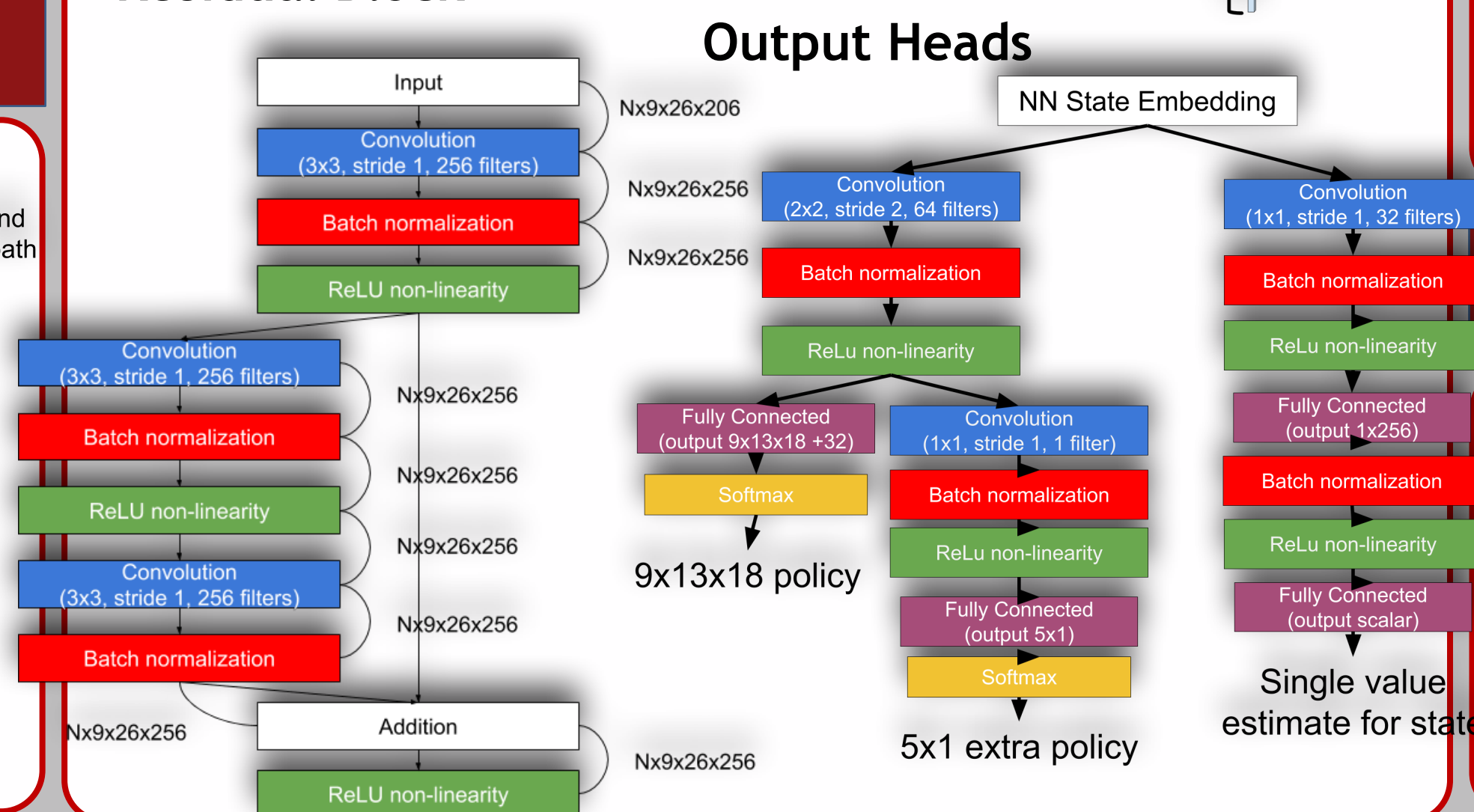$$(\boldsymbol{p}, v) = f_\theta(s) \text{ and } l = (z - v)^2 - \boldsymbol{\pi}^{\mathrm{T}} \log \boldsymbol{p} + c\|\theta\|^2$$

Neural Network Architecture

Multi-headed Architecture

Single Convolution

Residual Block 39x

Dual Policy Head

Value Head

206 256 256 64

Residual Block

Input Nx9x26x206
Convolution (3x3, stride 1, 256 filters)
Batch normalization Nx9x26x256
ReLU non-linearity Nx9x26x256
Convolution (3x3, stride 1, 256 filters)
Batch normalization Nx9x26x256
ReLU non-linearity Nx9x26x256
Convolution (3x3, stride 1, 256 filters)
Batch normalization Nx9x26x256
Addition Nx9x26x256
ReLU non-linearity

Output Heads

NN State Embedding

Convolution (2x2, stride 2, 64 filters)
Batch normalization
ReLu non-linearity
Fully Connected (output 9x13x18 +32)
Softmax
9x13x18 policy

Convolution (1x1, stride 1, 1 filter)
Batch normalization
ReLu non-linearity
Fully Connected (output 5x1)
Softmax
5x1 extra policy

Convolution (1x1, stride 1, 32 filters)
Batch normalization
ReLu non-linearity
Fully Connected (output 1x256)
Batch normalization
ReLu non-linearity
Fully Connected (output scalar)
Single value estimate for state

Results

Player Faction/AI Algorithm	Human Player	State of the Art AI	Alpha TM
Halflings	133.32	92.21	32.11
Engineers	127.72	77.12	34.12

Future Work

Introduce Supervised Learning to Stabilize Training
- Difficulty training using only self-play
- There are XX,000 games available online among human and existing AI player - use this information to jump-start self-play, similar to AlphaGo.

Expand to Multi-Player
- Current work focuses on 2-player games
- Next steps is to expand self-play to multi-player

Introduce Different NN Architectures
- Introduce better policy and value estimators
- Possible use for RNNs or LSTMs for long-term memory storage and strategy

References and Acknowledgements

[1] BoardGeek. Terra mistica: Statistics, 2011
[2] D. Silver, et. al, Mastering the game of go with deep neural networks and tree search. Nature, 529:484 EP –, Jan 2016. Article.
[3] D. Silver, et. al, Mastering chess and shogi by self-play with a general reinforcement learning algorithm. CoRR, abs/1712.01815, 2017.
[4] D. Silver, et.al. Mastering the game of go with- out human knowledge. Nature, 550:354 EP –, Oct 2017. Article.
[5] tmai. Terra mystic ai implementations, 2018