AlphaGo
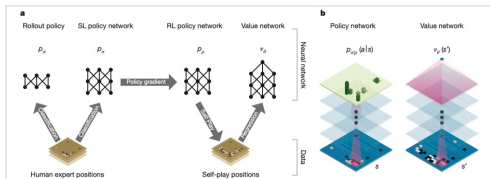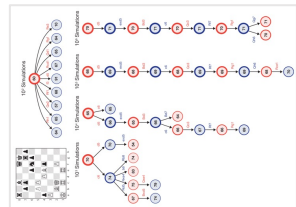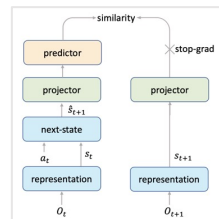Nature 2016
with Expert Data, Domain
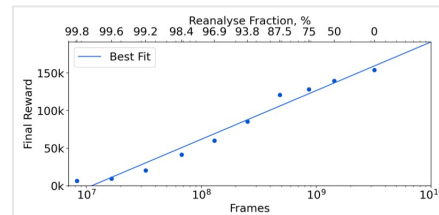Knowledge and Known Rules;
Value NN, Policy NN;
Env: Go

AlphaZero
Science 2018
with Known Rules;
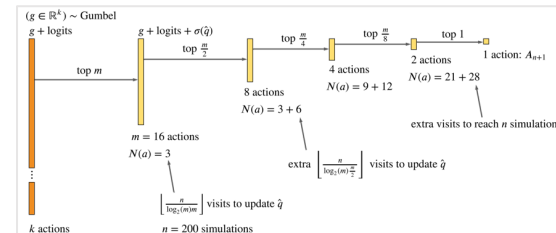Restructure Self-Play;
Env: Go, Chess, Shogi

EfficientZero
NeurIPS 2021
Self-Supervised Consistency Loss;
Value Prefix Prediction;
Model-Based Off-Policy Correction;
Env: Atari, DMC

MuZero Unplugged
NeurIPS 2021
Reanalyse;
Offline RL;
Env: Atari and DMC Dataset

Gumbel MuZero
ICLR 2022
Planning with Gumbel;
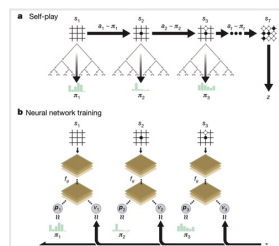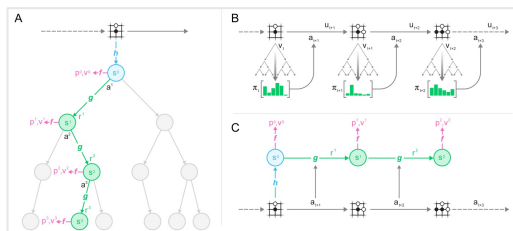Planning at Non-Root Nodes;
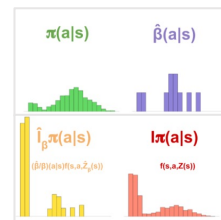Env: Go, Chess, Atari
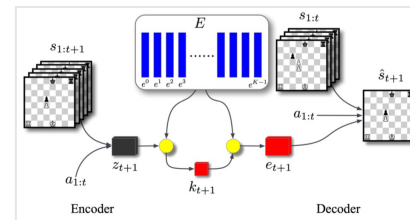
2016 2017 2018 2020 2021 2022

AlphaGo Zero
Nature 2017
with Known Rules;
A single Residual NN
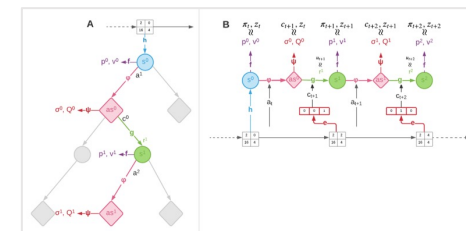for Value and Policy;
Self-Play;
Env: Go

MuZero
Nature 2020
without Any Knowledge;
Representation State;
Model-Based;
Env: Go, Chess, Shogi, Atari

Sampled MuZero
ICML 2021
Sample-Based Policy Iteration;
Extend to Complex Action Space;
Env: Go, DMC, RWRL

VQVAE MuZero
ICML 2021
VQ Model;
Stochastic MCTS;
Env: Chess Dataset, DM Lab

Stochastic MuZero
ICLR 2022
Stochastic Model;
Stochastic Tree Search;
Env: 2048, Backgammon, Go