EURECOM
Sophia Antipolis

GRADUATE SCHOOL AND RESEARCH CENTER
IN DIGITAL SCIENCE

PROJECT S6 GROUP 5

4 THE WIN

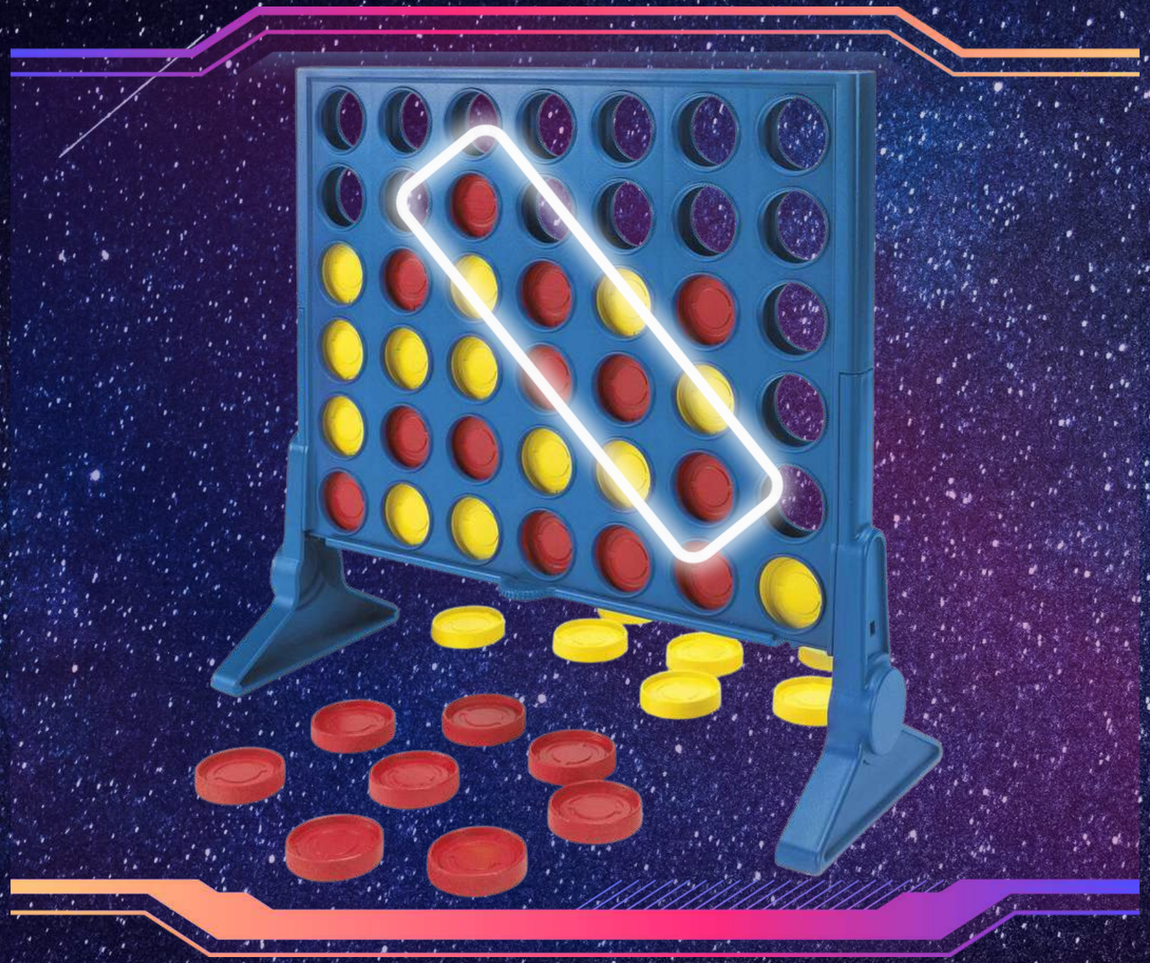
INSTITUT CARNOT
Télécom & Société numérique

INSTITUT CARNOT
Télécom & Société numérique

EURECOM

SUMMARY

HISTORICAL GAME



INTERFACE

ALPHAGOZERO

has beaton the GO the world champion Lee Sedol 4 games to 1 1

has beaton the GO The updated AlphaZero crushed Stockfish 8 in a new 1,000-game match, scoring +155 -6 =83the world champion Lee Sedol 4 games to 1 1

A technology powerfull enough to completely dominate human intelligence on the board

ALPHAGOZERO

A quick overview of a simple method

A Monte-Carlo Tree Search (MCTS)

A current_best model, playing against himself, generating data

Apprenticeship mode, learning from the data gathered

ALPHAGOZERO

A quick overview of a simple method

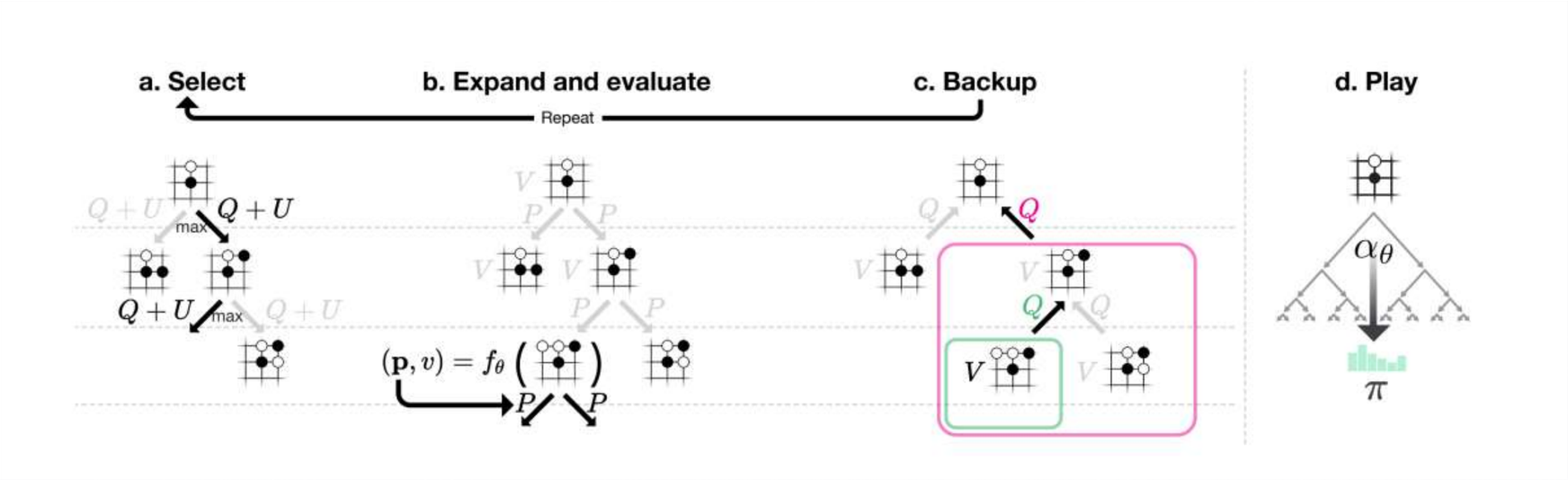
A Monte-Carlo Tree Search (MCTS)

A current_best model, playing against himself, generating data

Apprenticeship mode, learning from the data gathered

MONTE-CARLO TREE SEARCH

MCTS for AlphaGo Zero



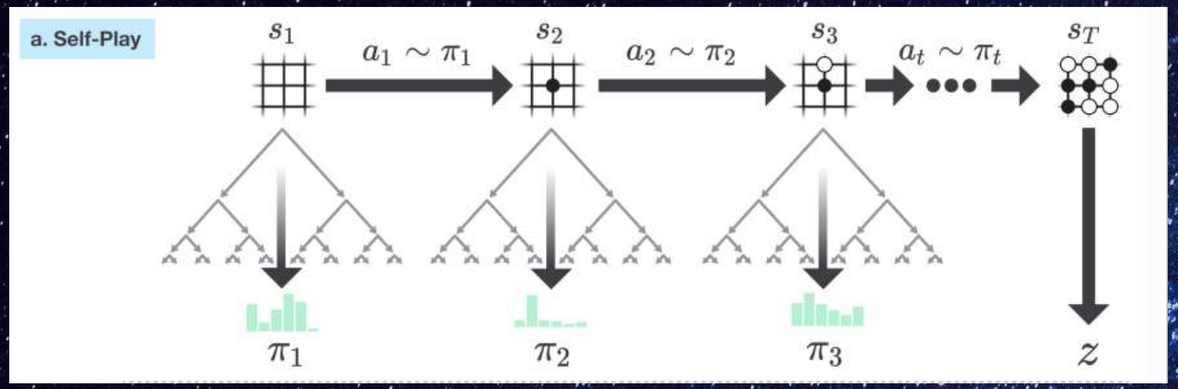
- Q(s,a): the expected reward for taking action a from state s, i.e. the Q values
- N(s,a): the number of times we took action a from state s across simulations
- P(s, ·)=P the initial estimate of taking an action from the state s according to the policy returned by the current neural network.

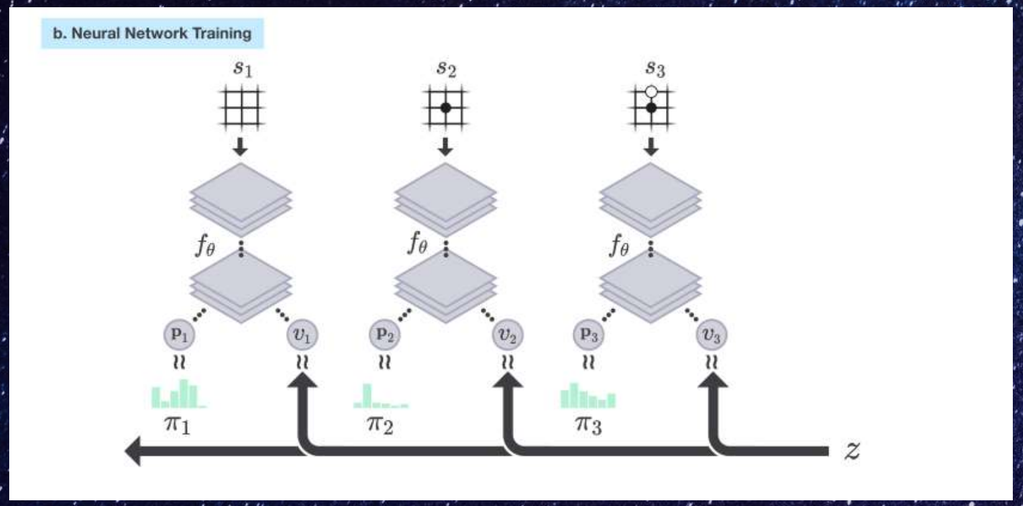
Where the actions are sampled following ;

$$U(s,a) \propto Q(s,a) + P(s,a)/1 + N(s,a)$$

SELF-PLAY AND TRAINING

Self-Play Reinforcement learning





The data for each time-step t is stored as a list of (st , πt , zt),

The Neural Network new parameters θi are trained from the uniformly sampled dataset, and adjust the loss function using gradient descent

$$l = (z - v)^2 - \boldsymbol{\pi}^\top \log \mathbf{p} + c||\theta||^2$$

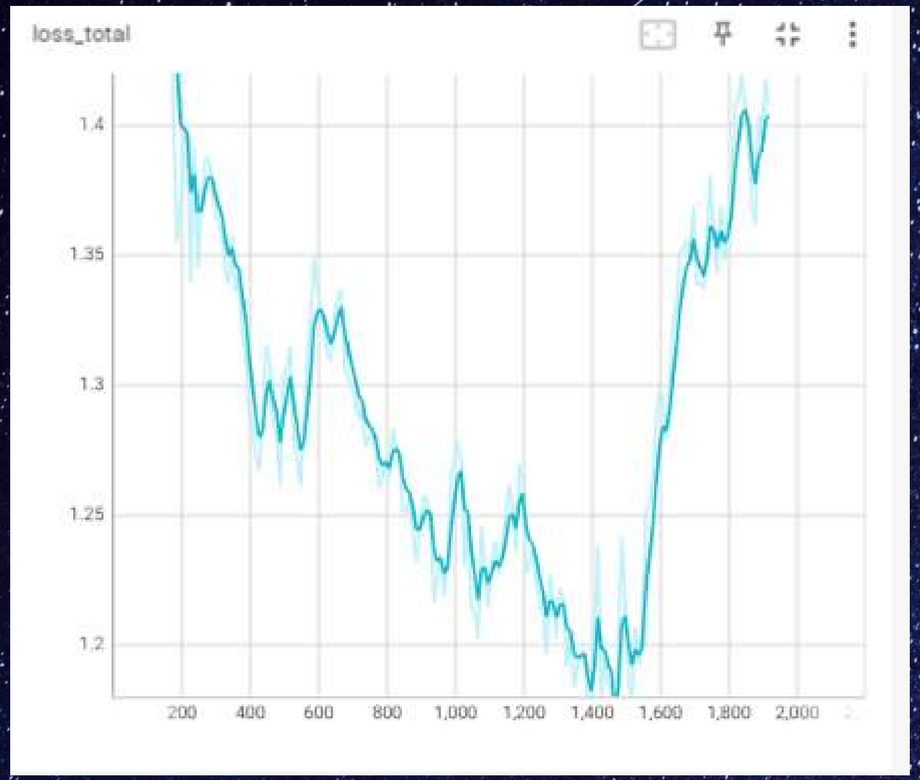
MSE between the actual winner and the network's prediction

cross-entropy losses, between P and π

CONNECT4 RESULTS

FIRST TRAINING (1.6HOUR, 2K GAMES)

LOSS FUNCTION OVER THE STEPS



SOME HYPERPARAMETERS ::

LEARNING_RATE = 0.1 # controls that is the step size of the adjustment of the parameters of the NN

BATCH_SIZE = 256 # specifies the number of training samples used in each iteration of gradient descent

CONNECT4 RESULTS

SECOND TRAINING (3.13 DAYS, 87,5K GAMES)

LOSS FUNCTION OVER THE STEPS

HYPERPARAMETER UPDATE

LEARNING_RATE = 0.1 # in the DeepMindPaper is adjusted

BATCH_SIZE = 512 # DeepMind Paper = 2048

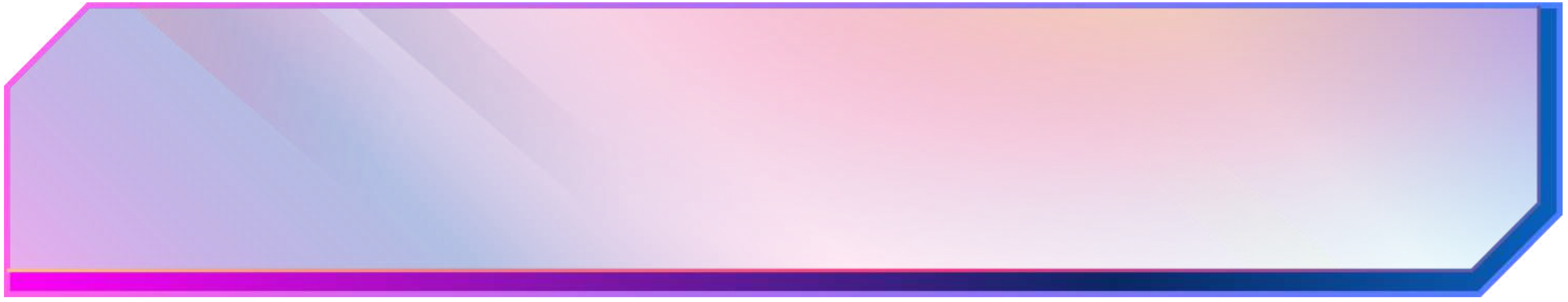
DECREASING THE LEARNING RATE OVER STEP TO MATCH DEEPMIND'S IMPLEMENTATION

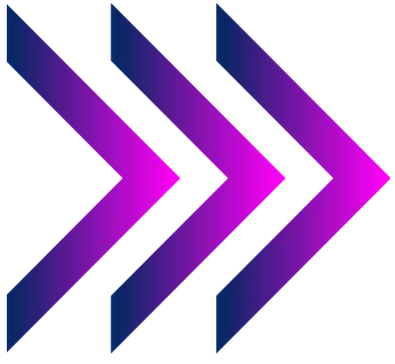
CONNECT4 RESULTS

FINAL MODEL AND COMMENT

- AN AI THAT IS SKILLED
- VERY ENJOYABLE TO PLAY AGAINST (PLAYS INSTANTLY AND HAS HUMAN-LIKE GAMEPLAY)
- TRAINING THAT PROVIDED 20 DIFFERENT MODELS, WHICH MADE THE CAMPAIGN IMPLEMENTATION POSSIBLE
- MAKES MISTAKES ;
  - COMPUTING THE AVERAGE OF DIFFERENT MODEL TO PREVENT
  - IMPLEMENTATION OF A FUNCTION THAT CHECKS LOSING POSITION

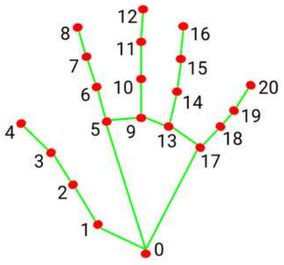
steps





Hand recognition





0. WRIST
1. THUMB_CMC
2. THUMB_MCP
3. THUMB_IP
4. THUMB_TIP
5. INDEX_FINGER_MCP
6. INDEX_FINGER_PIP
7. INDEX_FINGER_DIP
8. INDEX_FINGER_TIP
9. MIDDLE_FINGER_MCP
10. MIDDLE_FINGER_PIP
11. MIDDLE_FIN(
12. MIDDLE_FIN(
13. RING_FINGEF
14. RING_FINGEF
15. RING_FINGEF
16. RING_FINGEF
17. PINKY_MCP
18. PINKY_PIP
19. PINKY_DIP
20. PINKY_TIP

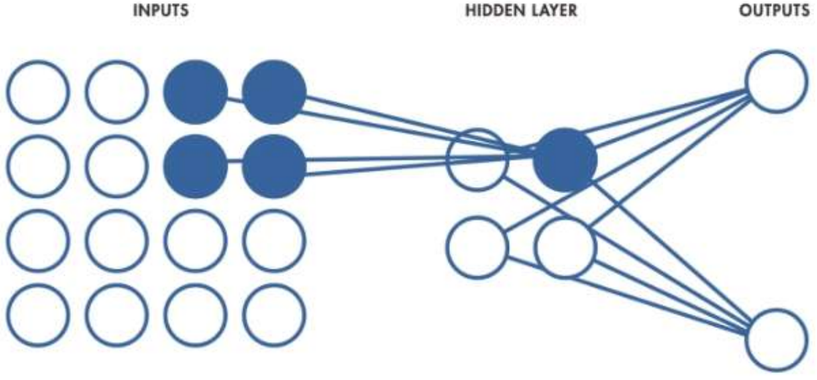
The order and labels of MediaPipe's Hands keypoints



Video Preprocessing video Hand tracking : landmark extraction and hand detection

Deep learning model : Convolutional Neural Networks (CNNs)

CONVOLUTIONAL
NEURAL
NETWORK



Wireless Communication