Reinforcement Learning in Action: Creating Arena Battle AI for 'Blade and Soul'

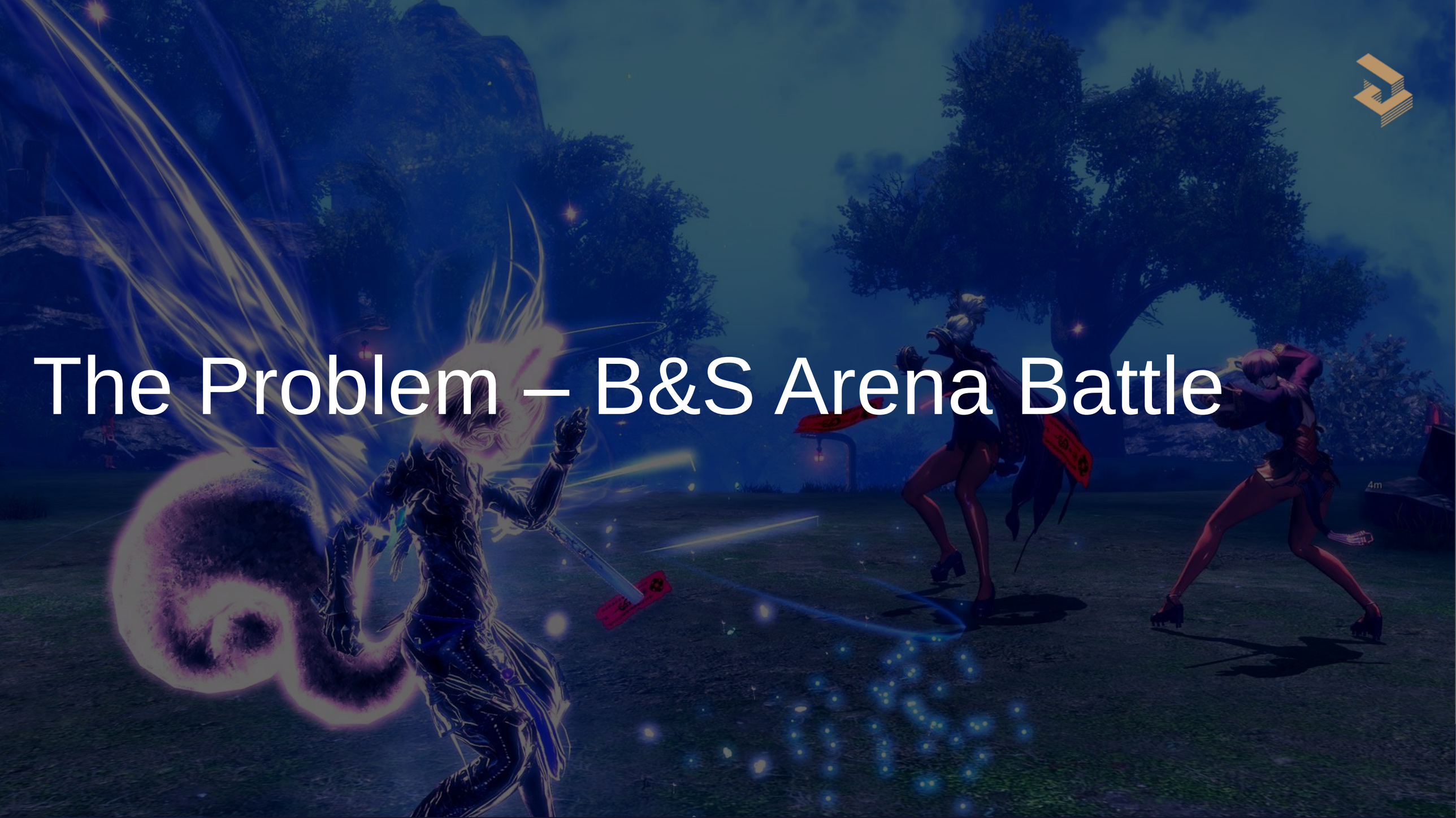
Jinyun Chung, Seungeun Rho
Reinforcement Learning Team / Game AI Lab / AI Center
NCSOFT

GAME DEVELOPERS CONFERENCE
MARCH 18–22, 2019 | #GDC19

Contents

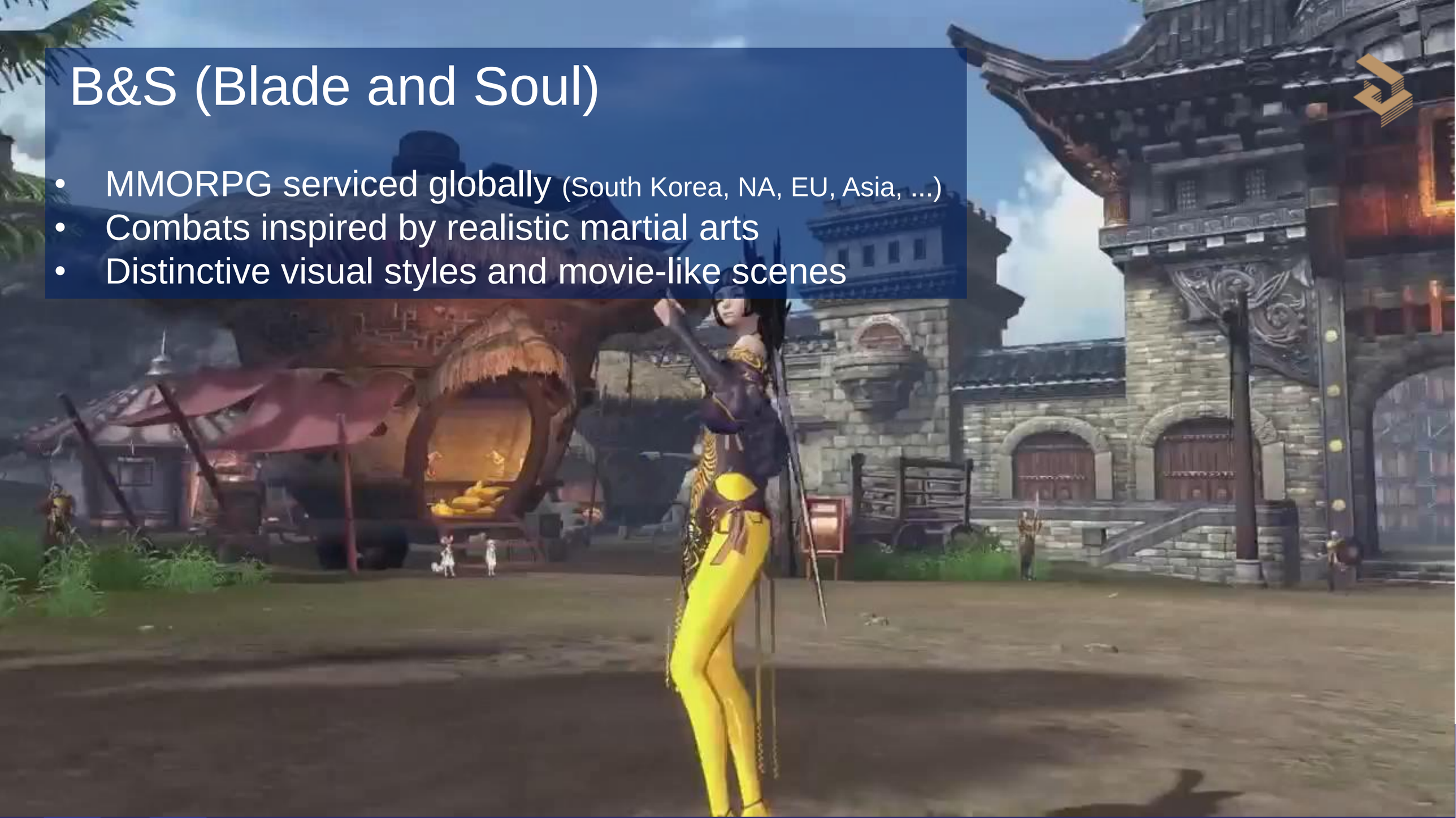
- The Problem – B&S Arena Battle
- Methods & Tips
- Experiments
- Demo & Blind Match

The Problem – B&S Arena Battle

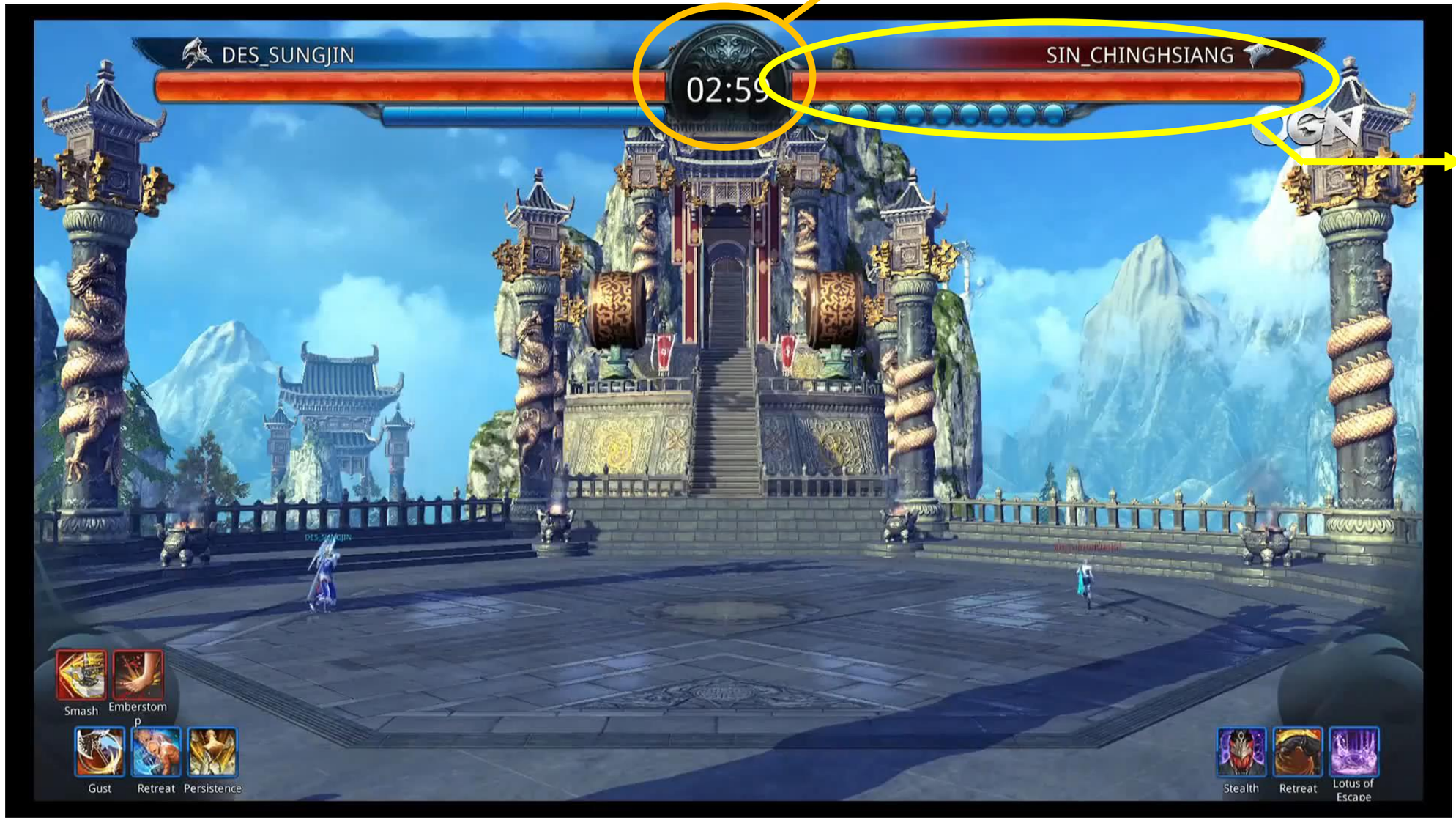


B&S (Blade and Soul)

- MMORPG serviced globally (South Korea, NA, EU, Asia, ...)
- Combats inspired by realistic martial arts
- Distinctive visual styles and movie-like scenes



B&S Arena Battle



Time limit
3 min

Goal
make opponent
HP to 0

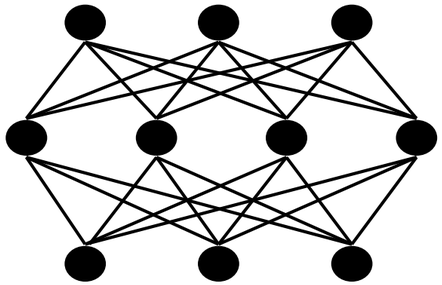
Research Goal

Pro-Level AI in 1v1 Arena Battle with RL

- AI with Hand-written rule

Simple…
Full of holes…

- RL-based AI of various styles



Various!
Flawless!

Research Goal

Pro-Level AI in 1v1 Arena Battle with RL

- With 3 different styles
  - Aggressive, Balanced, Defensive

- Conditions
  - Class : Destroyer vs Destroyer
  - Skill Setting : Equal

What does our agent do?

Agent

Observation

- HP, SP
- Distance
- Skill cool time
- …

Every 100 msec

Environment (Game)

Action

- Skill
- Move
- Target

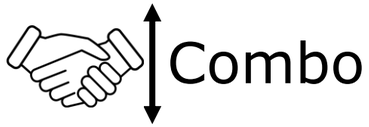
Skill Systems of B&S

Crowd Control

Drag

Ram

...



Combo

Counter

Damage Dealing

Judgement

Cleave

...

Resistance

High cost

Hurricane

Typhoon

...

Escape

High cost

Persistence

Retreat

...

Challenge 1) High Complexity

The game of Go

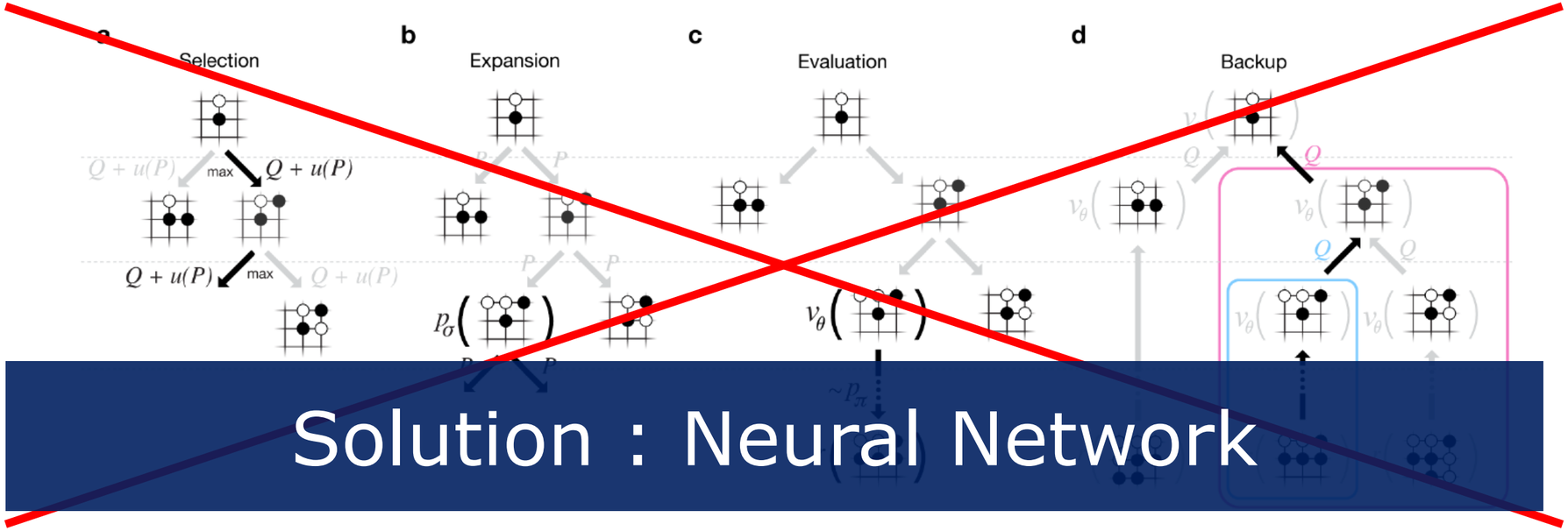
B&S Battle Arena

- Action Space ^ Game Length
  - Go ≈ 10^170
  - B&S ≈ 10^1800 = (10(skill) * 5(move) * 2(target)) ^ 900 (game length)

Solution : Action Space Reduction

Challenge 2) Real Time Response

- Decision Time Limitation
  - Go : 60 sec
  - B&S : 0.1 sec
- No Search Based Method!



Solution : Neural Network

Challenge 3) Generalization

- Best response differs for each player
  - Pro-gamers' play styles are diverse.
  - We don't know who will be the opponent
  - The agent should be generalized to response optimally against unknown players

Who will be the opponent?

Solution:
Opponent Pool of Various Selves

Challenge 4) Guiding Fighting Style

- 3 different fighting styles
  - Aggressive, Balanced, Defensive
- w/o hand-written rule

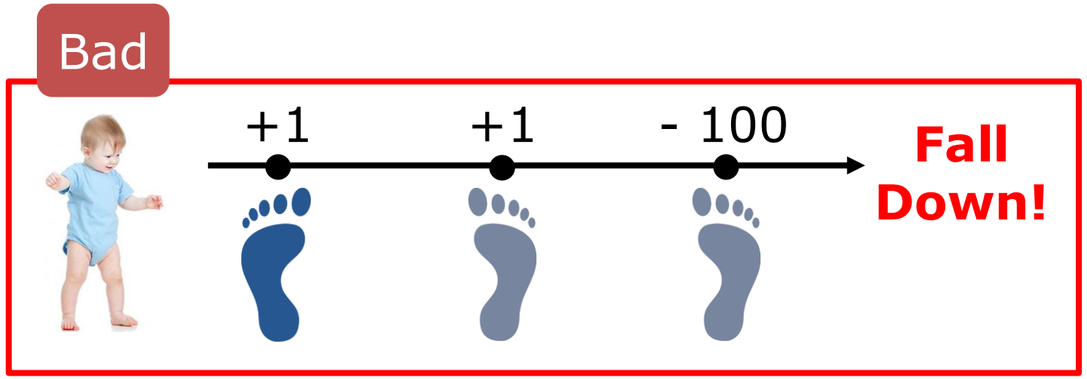
Solution: Reward Shaping

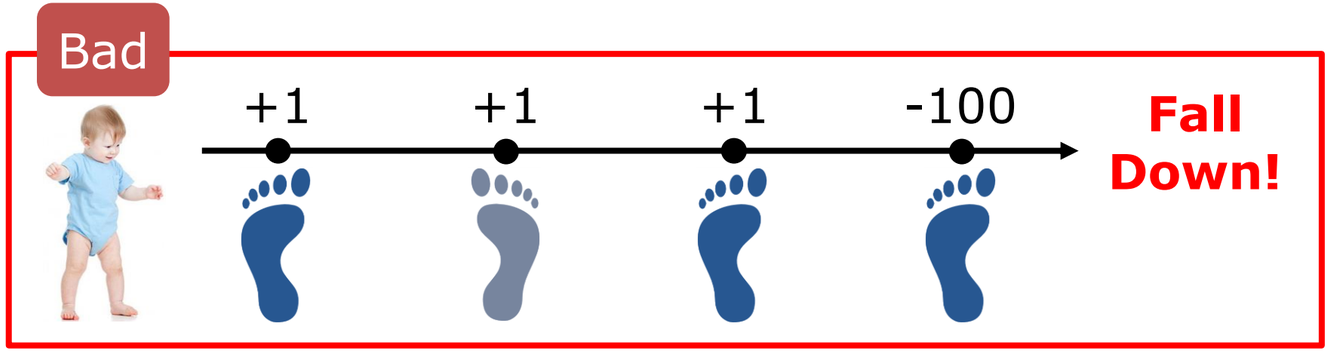
Methods & Tips

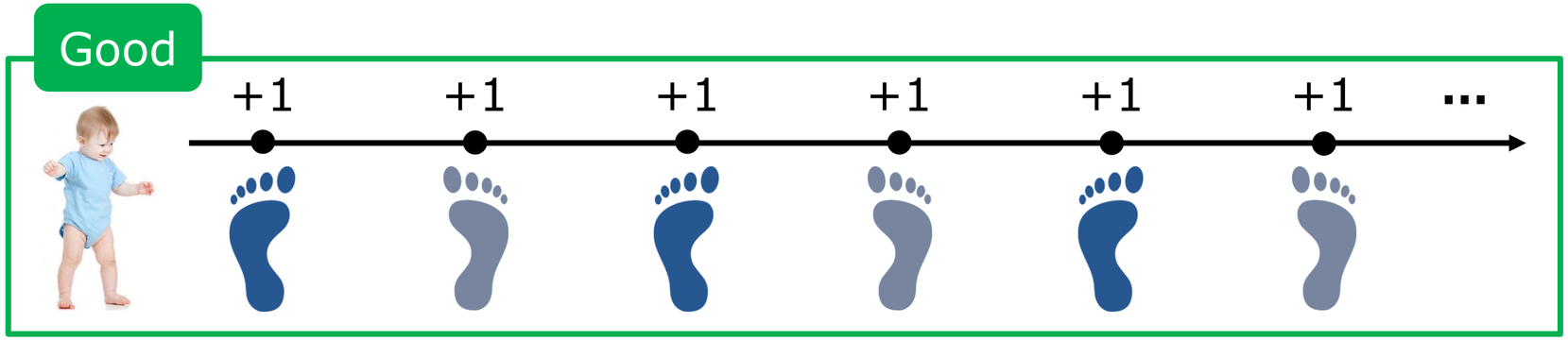
1. Basic Concepts of RL
2. Learning Process (features and rewards)
3. Engineering Techniques

Learning how to walk

image ref : https://www.bundoo.com/articles/signs-your-baby-is-ready-to-walk/







- Reward
  1. Every timestep : +1
  2. Fall down : -100
- Objective
  Maximizing cumulative reward
- Agent modify policy
  1. Reinforce actions in good sequences
  2. Suppress actions in bad sequences

DES_SUNGJIN
KR) GC BUSAN Red
1
End
3rd Place Match
Set 2
0
TW) PengGanDi
SIN_CHINGHSIANG
DES_SUNGJIN
SIN_CHINGHSIANG
Victory
Reward
16 Hit Combo
Cleave
Smash
Cleave
Wrath
Cleave
42
35
Smash
Emberstom
p
32
Gust
Retreat
Persistence
Stealth
Retreat
Lotus of
Escape

Additional Reward for Guiding Battle Style

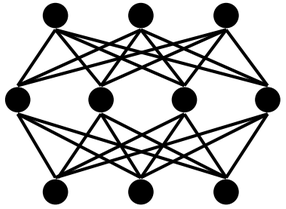
Basic

- **Win / Lose**
- **HP Gap**

Style

	Aggresive	Balanced	Defensive
HP Ratio (Me : Opponent)	4:6	5:5	6:4
Tick Penalty	-0.008	-0.004	0.0
Distance Penalty	-0.002		0
		Balanced,	Defensive

Agent-Environment Plot



Agent

Observation

(=Feature)

Represents how the agent sees the world

Reward

Every 100 ms

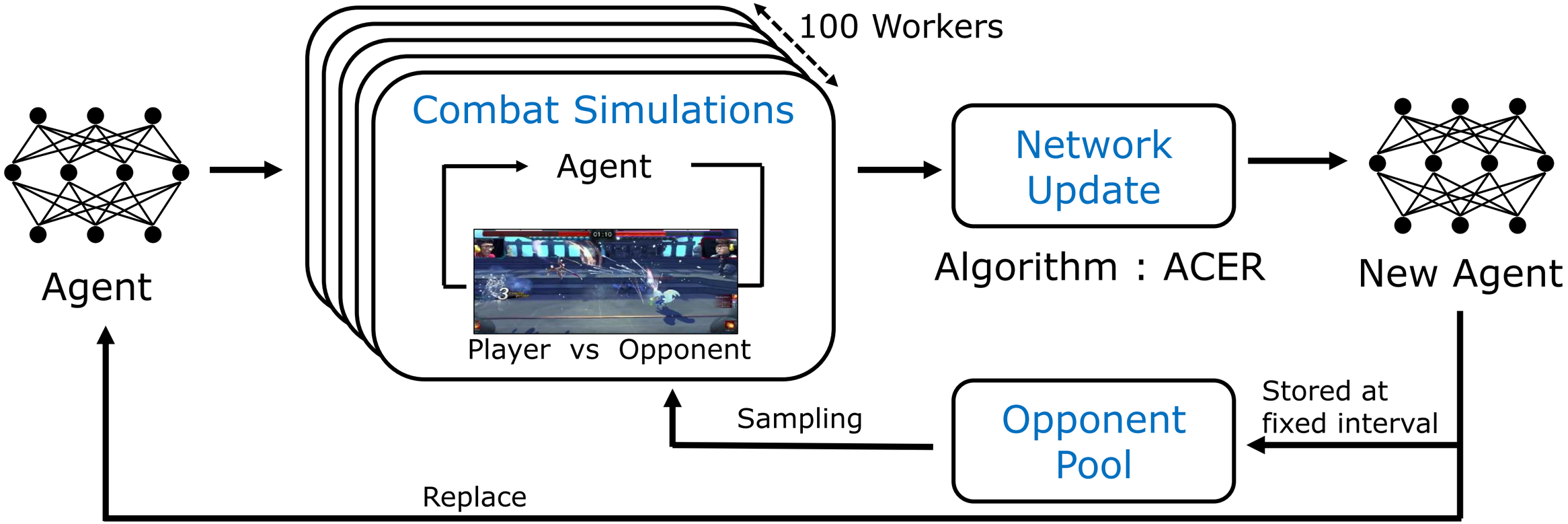
Action

(=Skill, Move)

Environment

(Game Simulator)

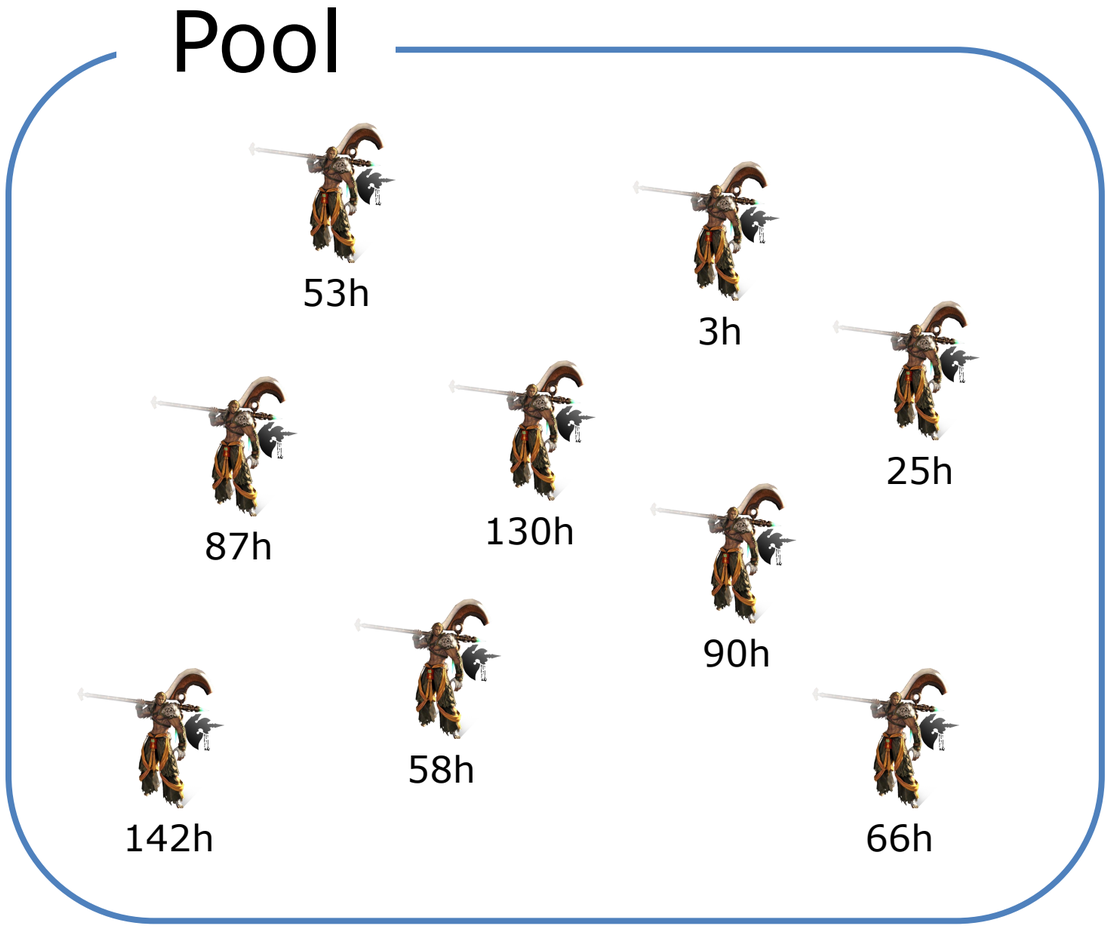
Learning Process



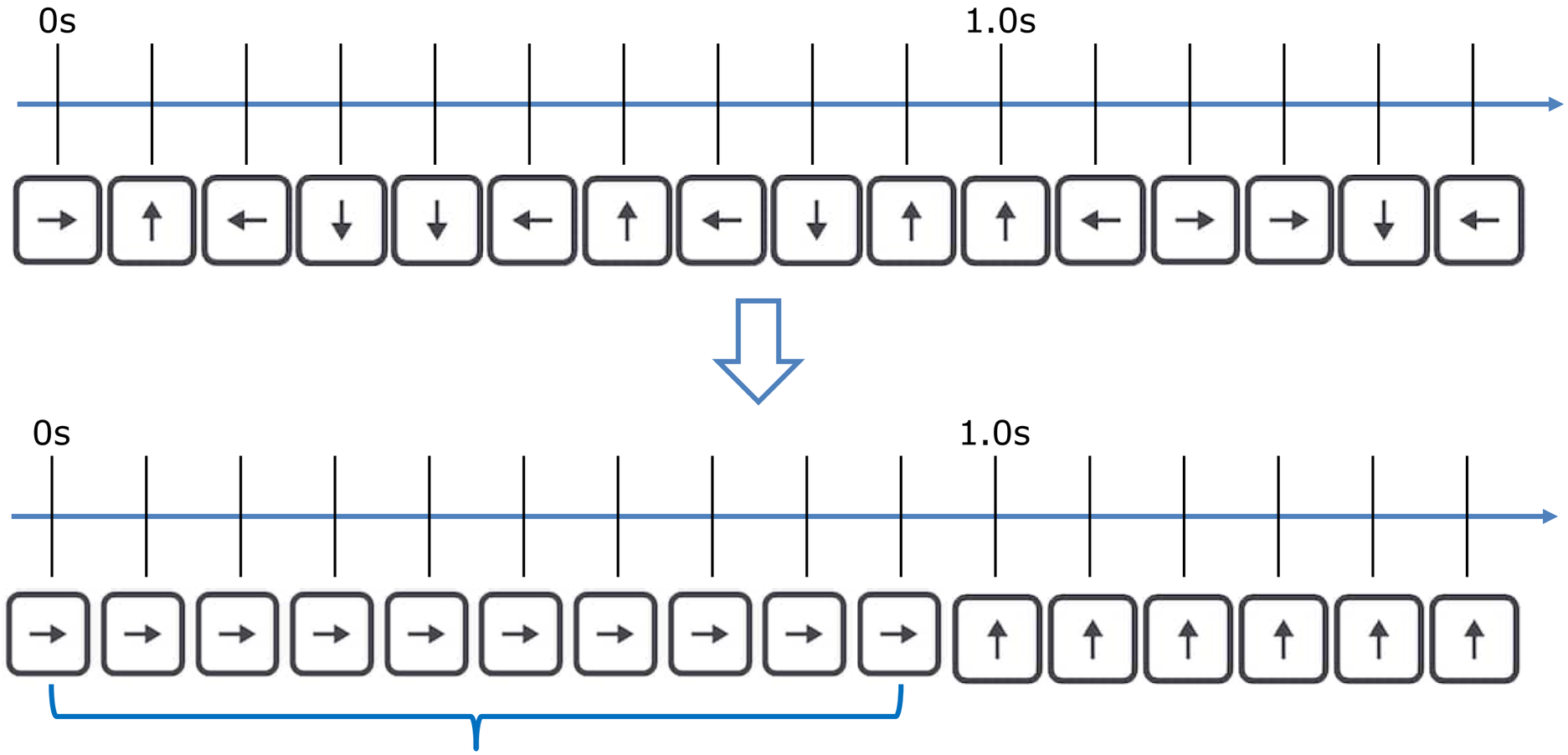
Who is the opponent? – The pool of past selves !

Latest,
163h learned

VS

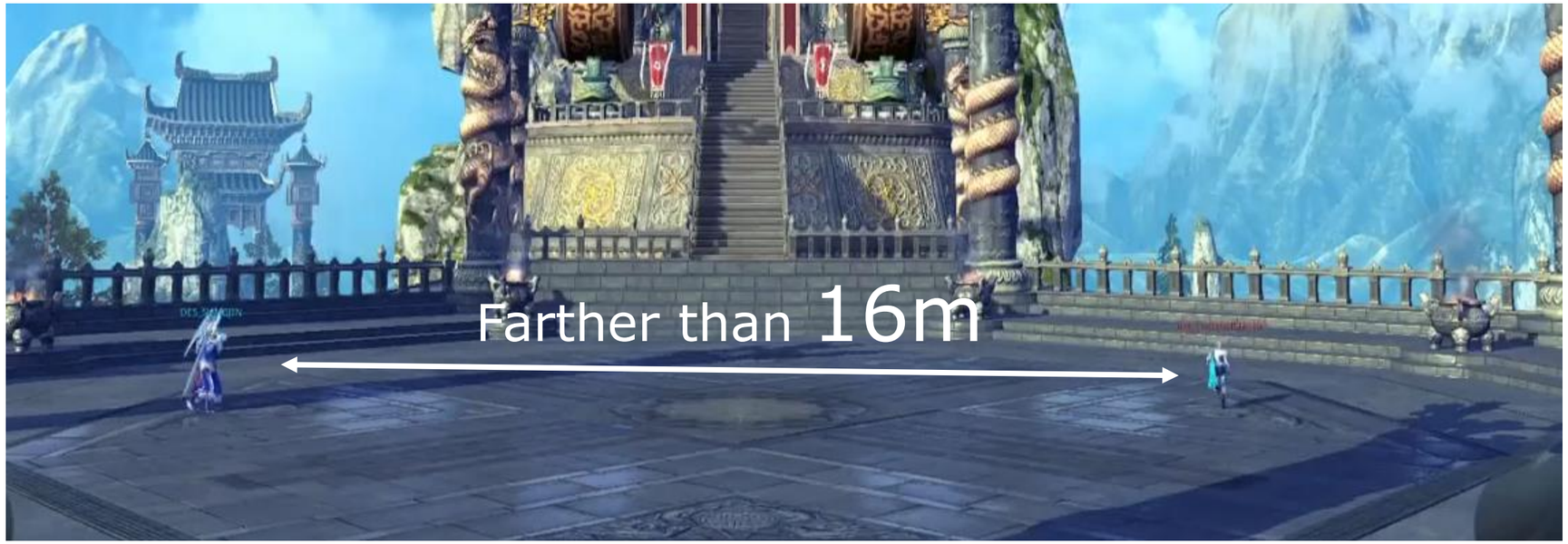


Move Space Reduction – Decision Frequency



Maintain Decision for 1 sec

Skill Space Reduction



- When the opponent is too far, most of skills are blocked
- Boosts the learning process in early phase

Feature Engineering

DRAG

- Discretization of continuous features
- Helps to distinguish fine skill range

<Distance feature>

0~1

⇩

0~3m	→	[1,0,0,0,0]
3~5m	→	[0,1,0,0,0]
5~8m	→	[0,0,1,0,0]
8~16m	→	[0,0,0,1,0]
16m~	→	[0,0,0,0,1]

Experiments

Sparring Partner – Built-in AI

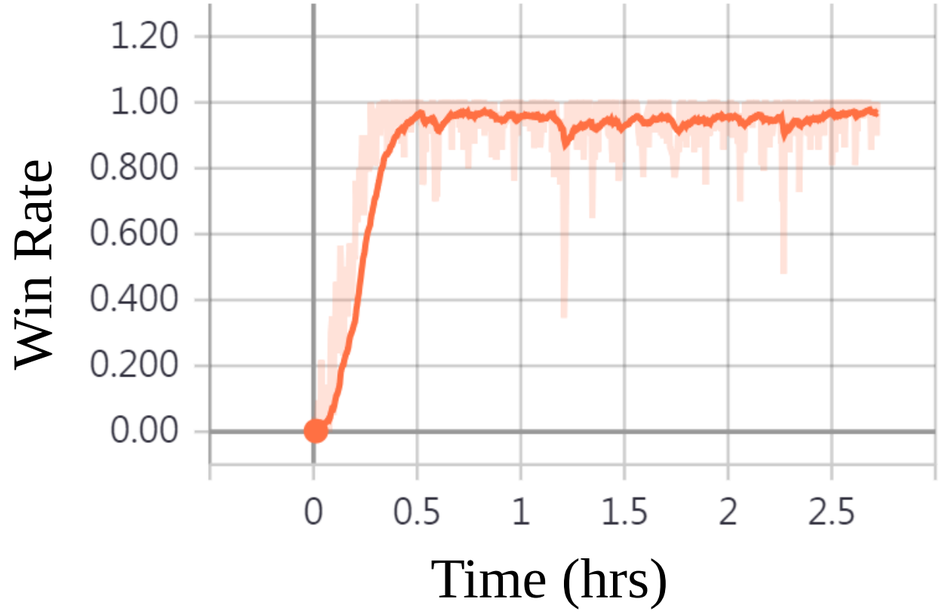
VS

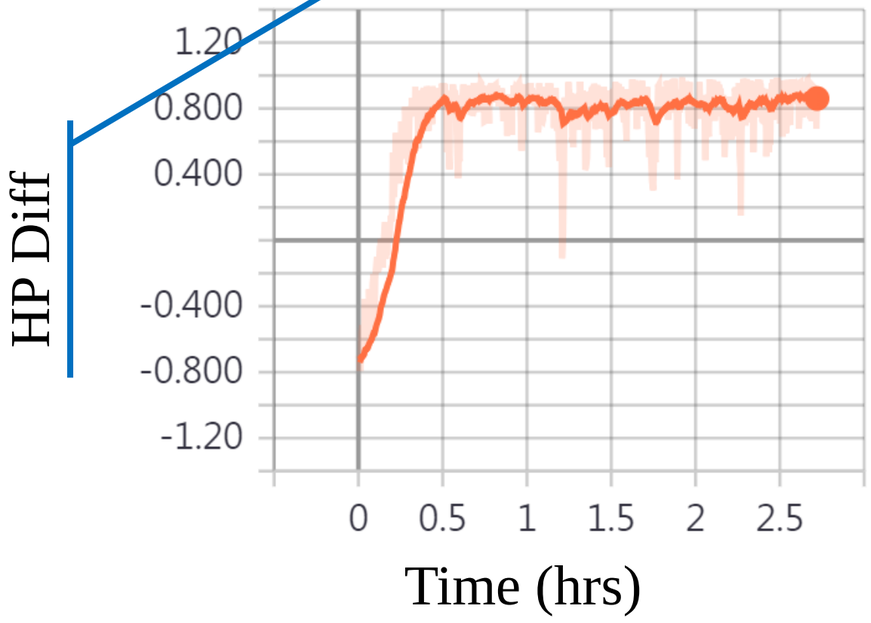
Built-in AI

(from Infinity Tower)

Sparring Partner – Built-in AI

$$HP_{me} - HP_{opponent}$$





- 90% win rate in 30 min
- 100% win rate in 5 hours

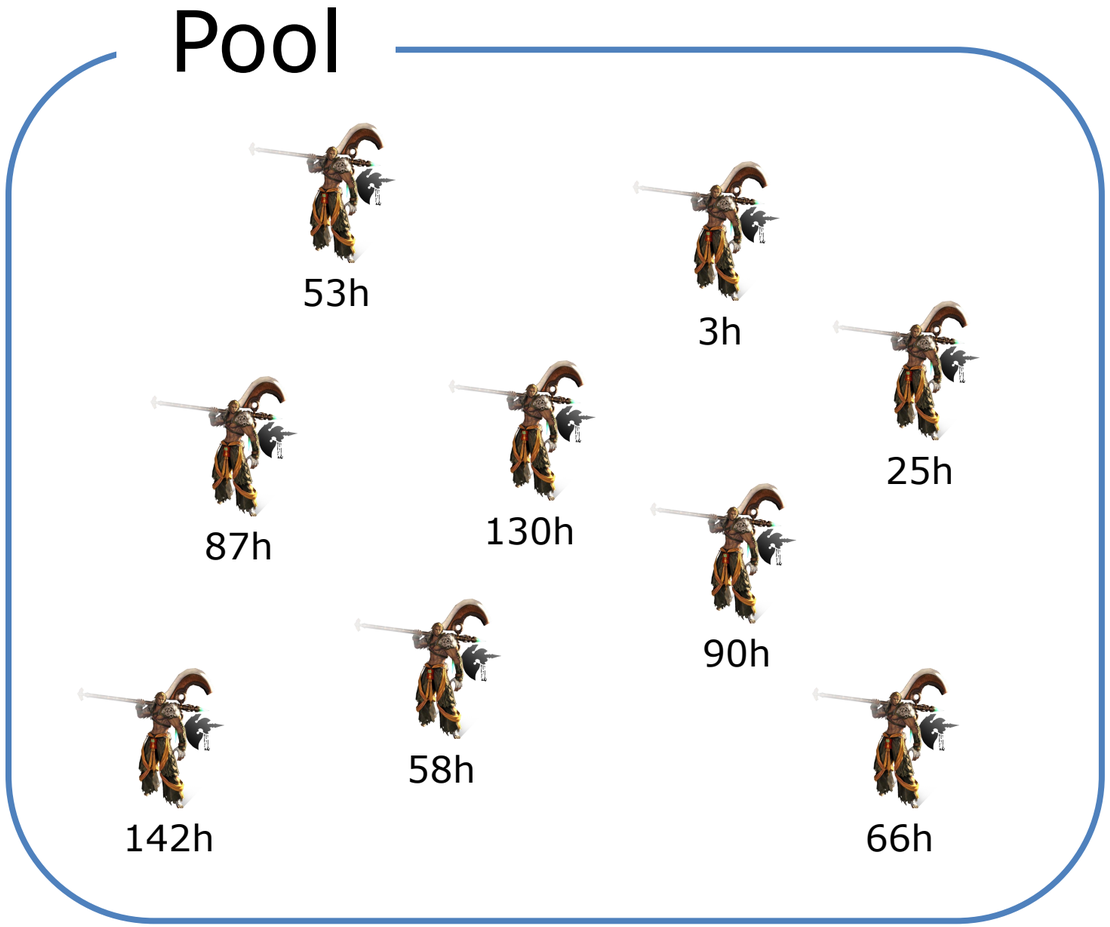
Level of difficulity against human :

Easy

Sparring Partner – Self play

Latest,
163h learned

VS



Level of difficulity against human :

Hard

Sparring Partner – Self play

Measuring generalization performance

- vs Built-in AI (Infinity Tower AI)
  - Training time 1 day : 84% win rate
  - Training time 2 days : 96% win rate

- vs Skillful human player (rating: top 10%)
  - 5 matches results : AI 5 - 0 Human
  - However … after about a hundred games of play, human adapt and exploit weak point of AI.
  - Training longer, weak points disappeared

Demo & Blind Match

Learning Progress

Initial

124 sec

1 day

42 sec

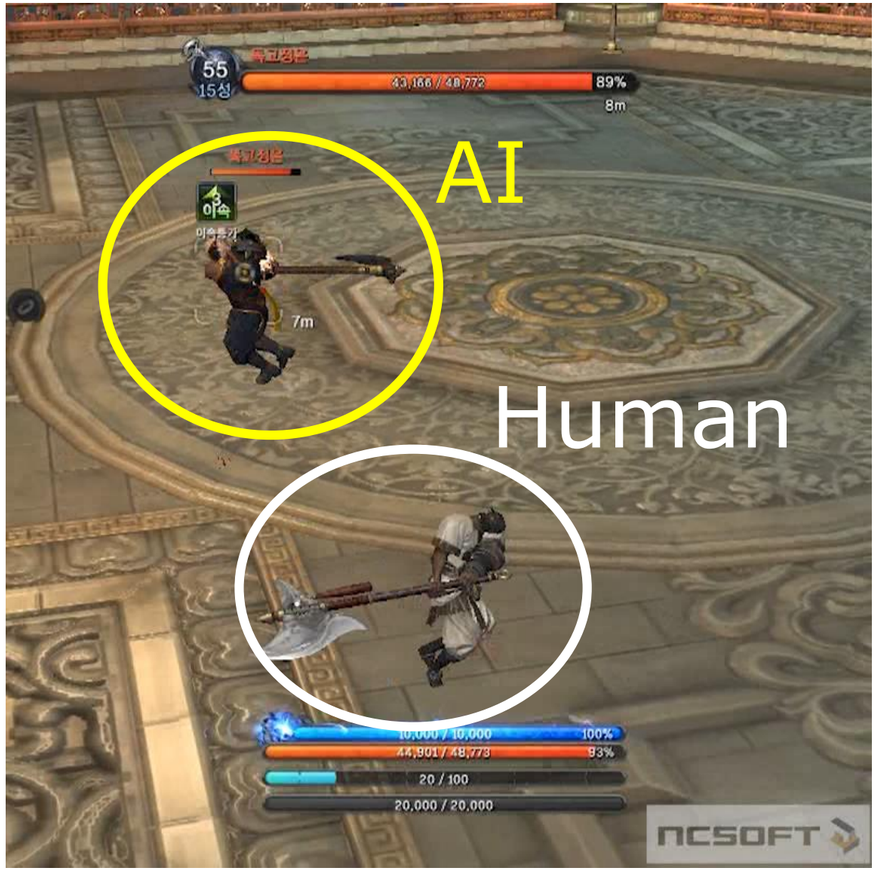
1 week

20 sec

Time consumption for reduce dummy's HP from **full** to **0**

Different Fighting Styles

Aggressive



Defensive

Pretest with professional gamers

AI - Human

	Aggressive	Balanced	Defensive
Pro-gamer 1	5-1	2-1	1-2
Pro-gamer 2	4-0	2-4	4-1
Total	9-1 (90%)	4-5 (44%)	5-3 (63%)

Pro-gamer 1 : Yuntae Son (The winner of 2017 B&S World Championship)
Pro-gamer 2 : Shingyeom Kim (The winner of 2015 and 2016 B&S World Championship)

Reinforcement Learning in Action: Creating Arena Battle AI for 'Blade and Soul'

Reinforcement Learning Team
Game AI Lab, AI Center
NCSOFT

Results

AI - Human

	Aggressive	Balanced	Defensive
Pro-gamer 1	5-1	2-1	1-2
Pro-gamer 2	4-0	2-4	4-1
Blind Match	**2-0**	**1-2**	**0-2**
Total	11-1 (92%)	5-7 (42%)	5-5 (50%)

Conclusion

- RL actually works!
- Can guide fighting styles w/o hand-written rules
- Pool of opponents help generalization
- Space reduction helps training

NCSOFT
Reinforcement Learning Team
Game AI Lab, AI Center
Jinyun Chung, jchung2050@ncsoft.com
Seungeun Rho, gloomymonday@ncsoft.com