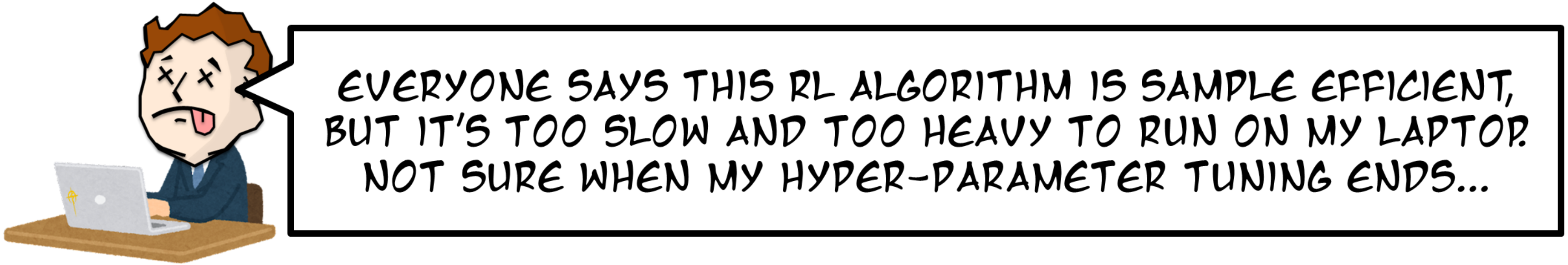
Dropout Q-Functions for Doubly Efficient Reinforcement Learning

Takuya Hiraoka [1,2], Takahisa Imagawa [2], Taisei Hashimoto [2,3], Takashi Onishi [1,2], and Yoshimasa Tsuruoka [2,3]

[1] NEC Corporation, [2] AIST, [3] The University of Tokyo

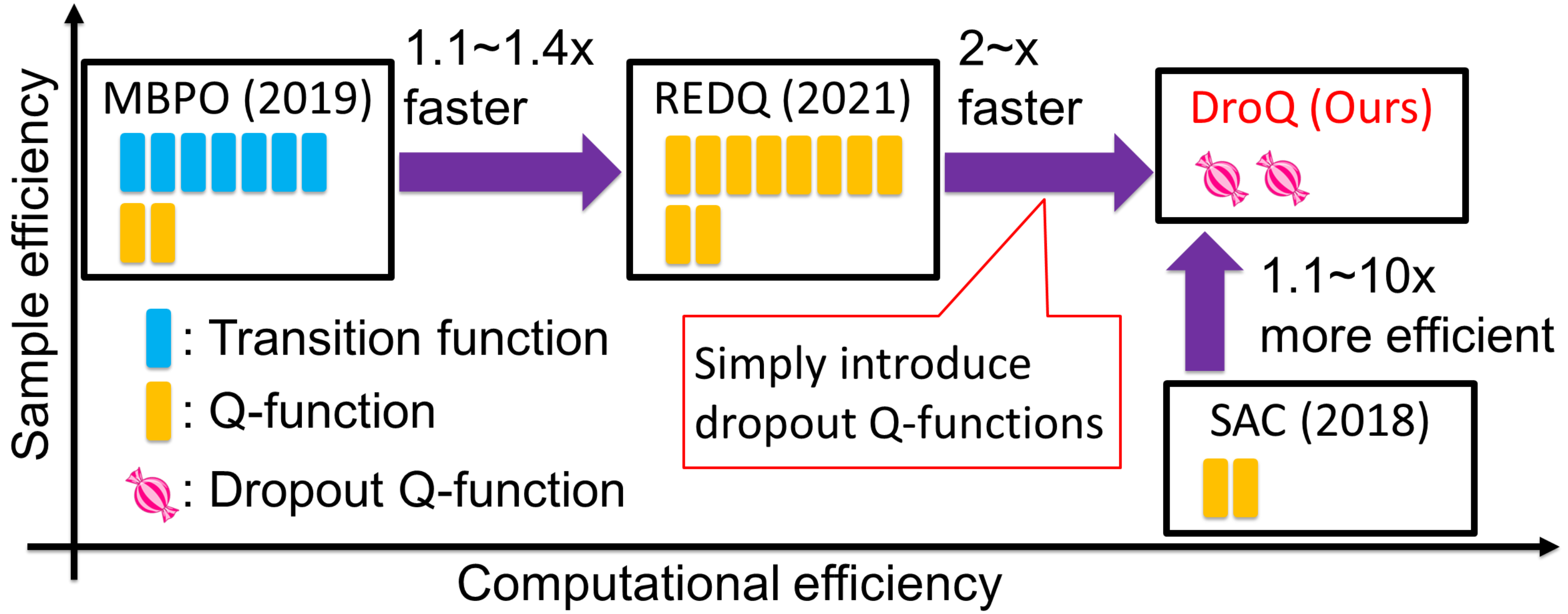
Introduction

- In general, **doubly efficient** RL methods that are not only **sample efficient** but also **computational efficient** are preferable.



Introduction (Contd.)

- We propose DroQ, a simple but doubly efficient RL method, by introducing dropout Q-functions (🍬) to REDQ.

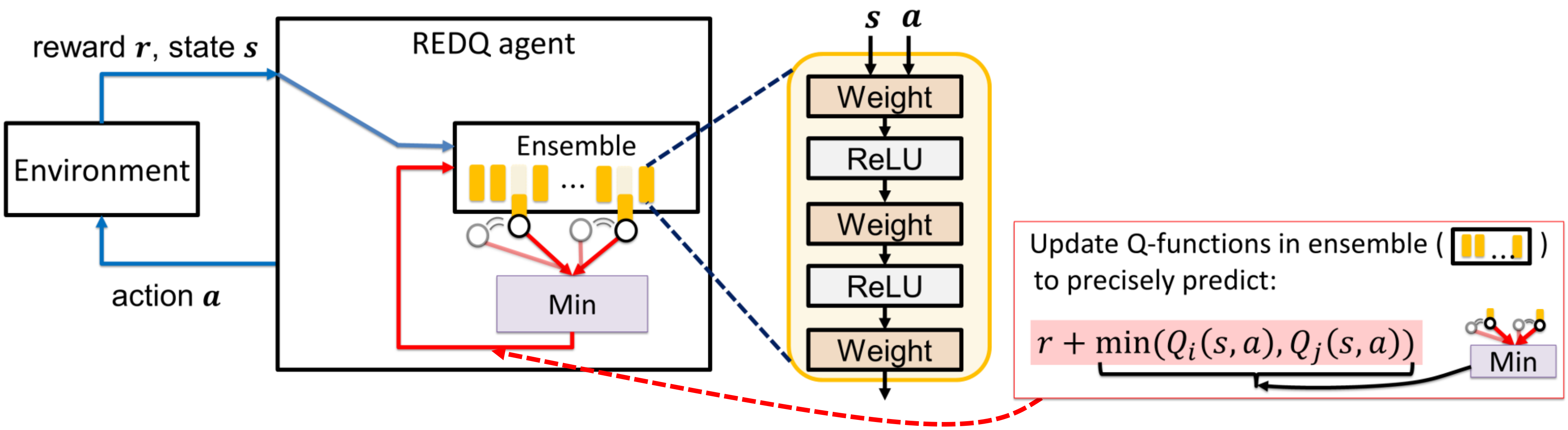


REDQ: Randomized Ensembled Double Q

NEC-AIST
AI Cooperative
Research Laboratory

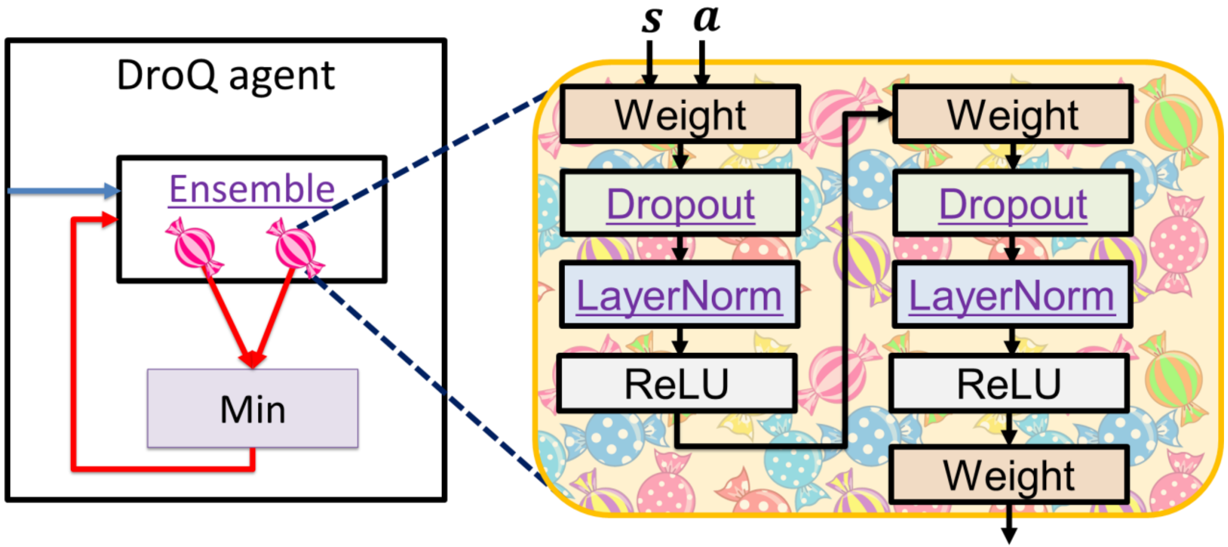
REDQ (Chen, 2021) is a sample-efficient RL method equipped with

- **High update-to-data (UTD) ratio:** number of Q updates (→) per environment interaction (→) is high (e.g., 20 updates per interaction).
- **Randomized ensemble:** a randomly selected subset ( ) of ensemble (...) is used at the target (Min) in the Q update (→).



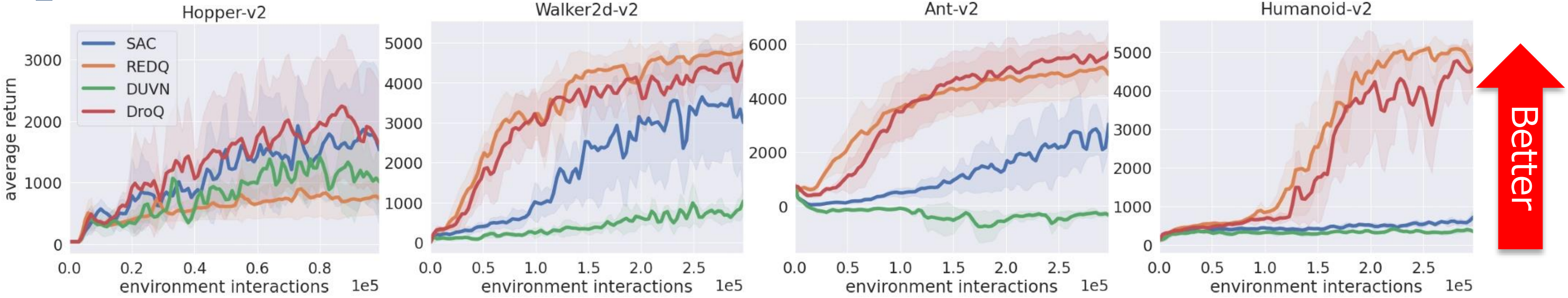
DroQ, the proposed method

- DroQ is a REDQ variant using a small ensemble of dropout Q-functions ( ) in which dropout (Dropout) and layer normalization (LayerNorm) are used.

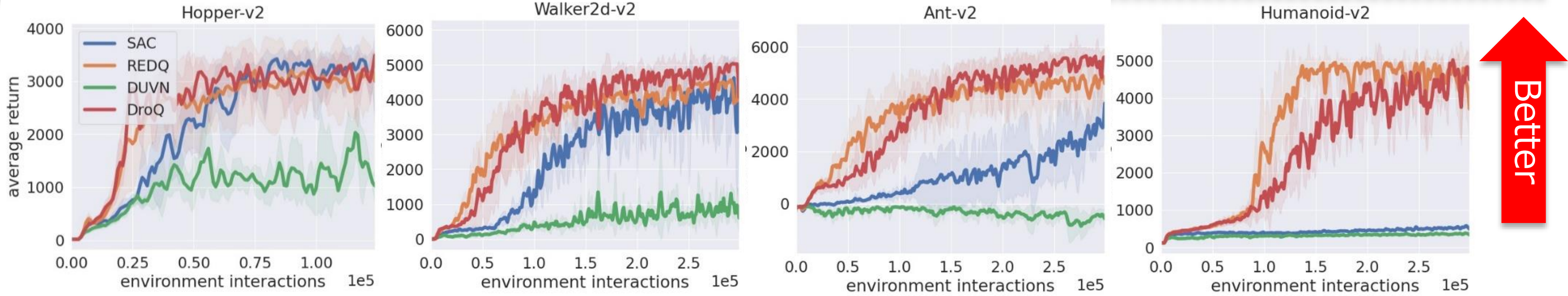


Q. How sample-efficient is DroQ ( )?

A. Better than SAC ( ) and almost the same as REDQ ( ).



Experiment on codebase A

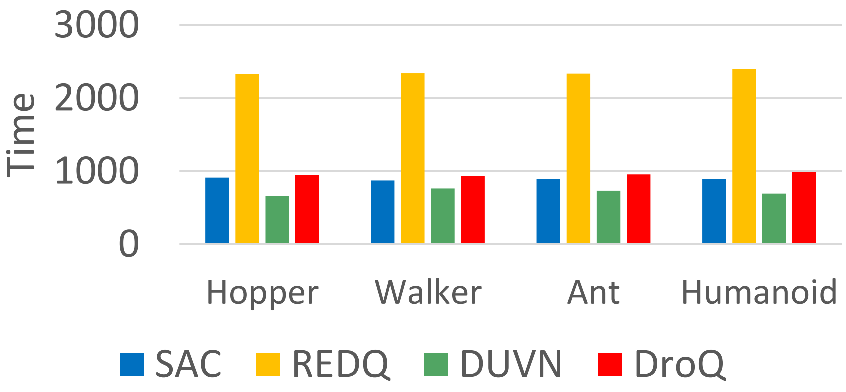


Experiment on codebase B

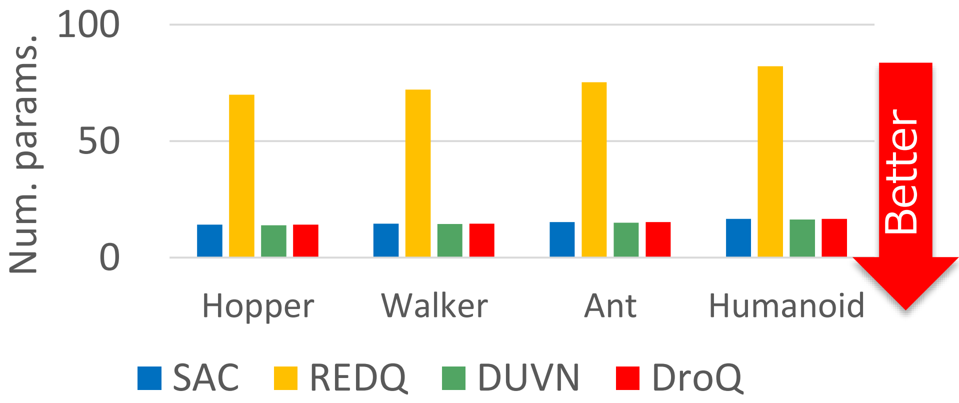
Q. How computationally efficient is DroQ?

NEC-AIST
AI Cooperative
Research Laboratory

A. Much better than REDQ and almost the same as SAC.



Times per 20 updates + 1 interaction (in msec)

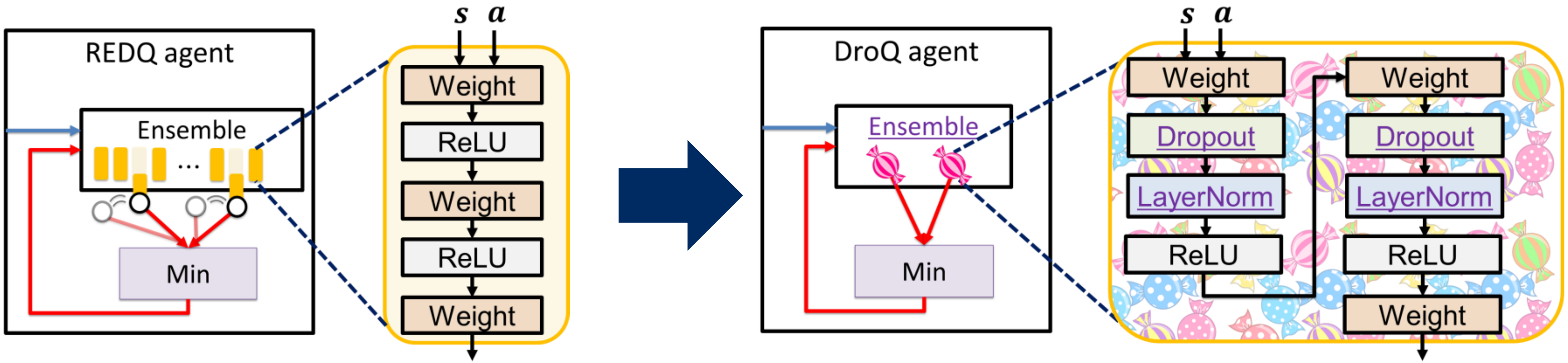


Number of parameters (/1e4)

Conclusion

NEC-AIST AI Cooperative Research Laboratory

- DroQ (REDQ + ) is simple but doubly efficient.



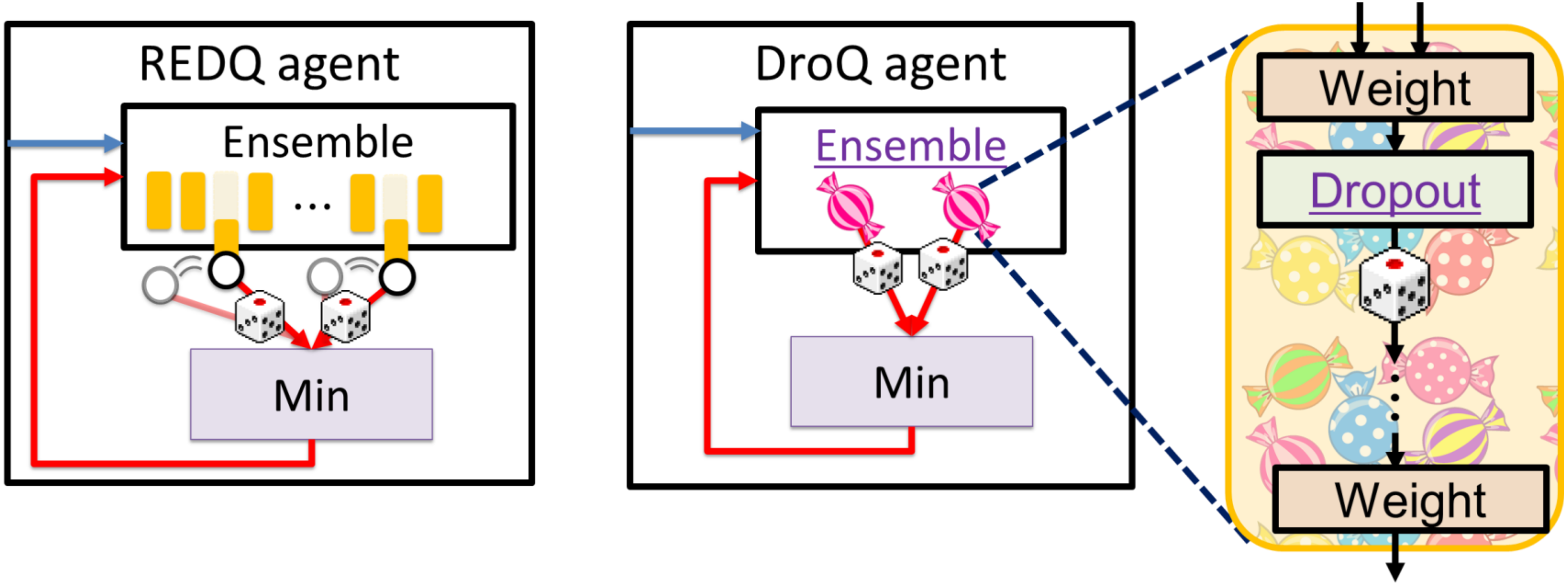
- **Are you interested in our work?**
Or feel that all we did was just randomly changing modules of the existing RL method?

INTRODUCTION
TO
OUR POSTER

Q. Why is dropout (Dropout) needed?

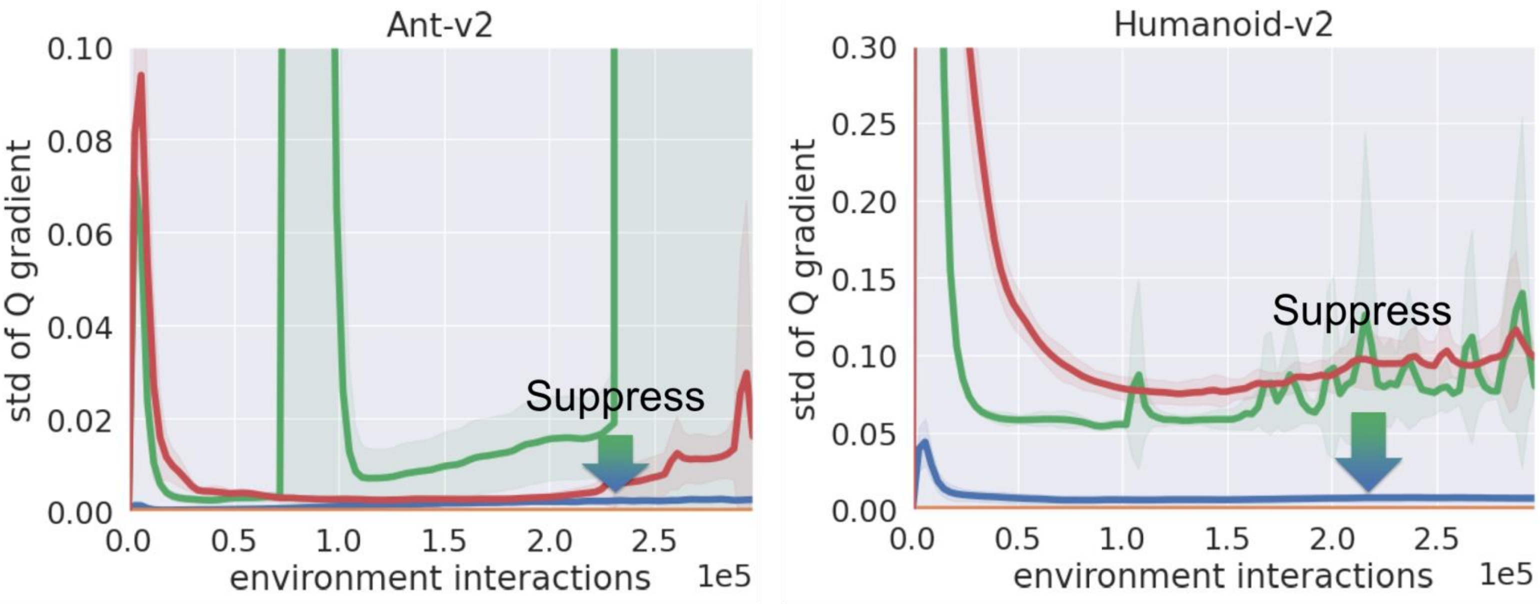
NEC-AIST AI Cooperative Research Laboratory

- A. To inject Q-function uncertainty ( ) to the target (Min), similarly to REDQ.



Q. Why is layer normalization (LayerNorm) needed?

A. To suppress ( ) the learning instability caused by dropout.

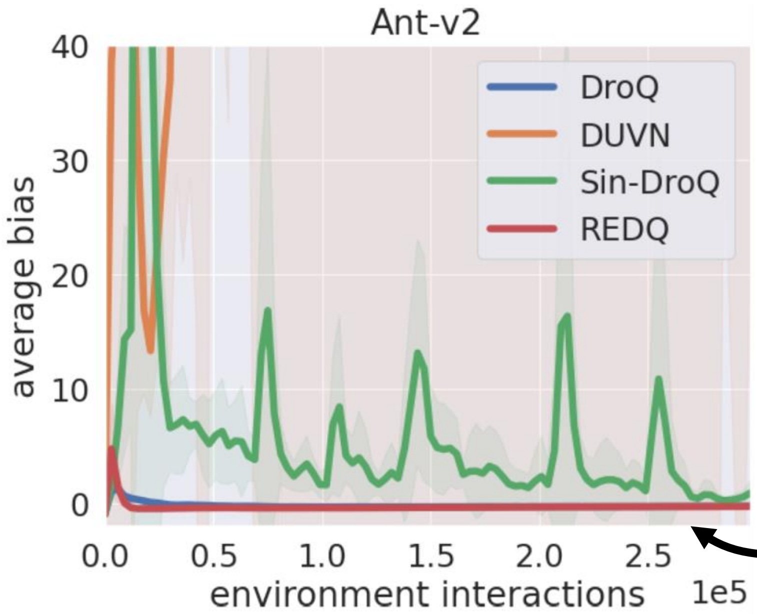


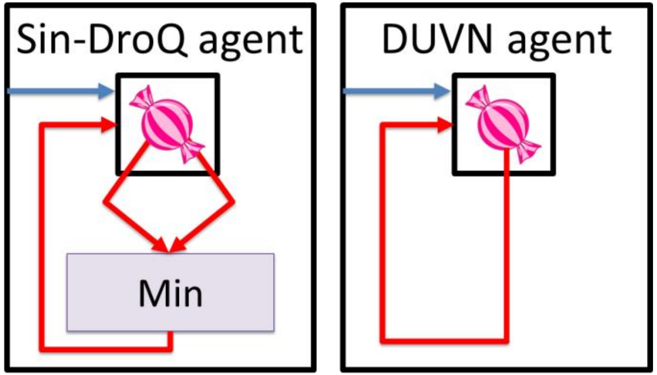
The standard deviation of the gradient of Q-loss w.r.t. parameters

DroQ — w/o Dropout — w/o LayerNorm — w/o Dropout nor LayerNorm

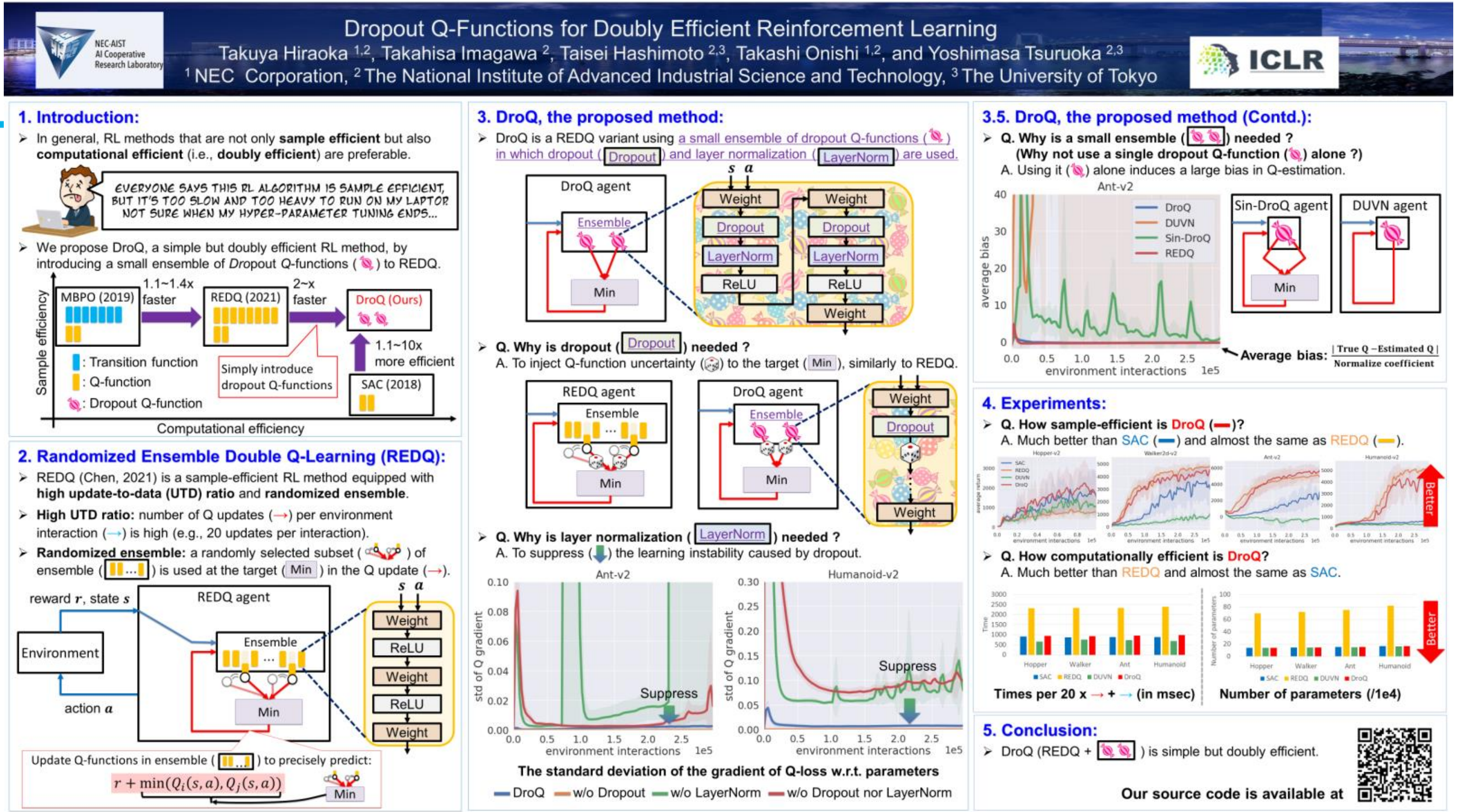
Q. Why is a small ensemble () needed?

- A. Using a single dropout Q-function () alone induces a large bias in Q-estimation.





Q-estimates bias calculated as $\dfrac{|\,\text{True Q} - \text{Estimated Q}\,|}{\text{Normalize coefficient}}$



Thank you for watching this video!

NEC-AIST
AI Cooperative
Research Laboratory