MITSUBISHI ELECTRIC
Changes for the Better

ICML | 2020

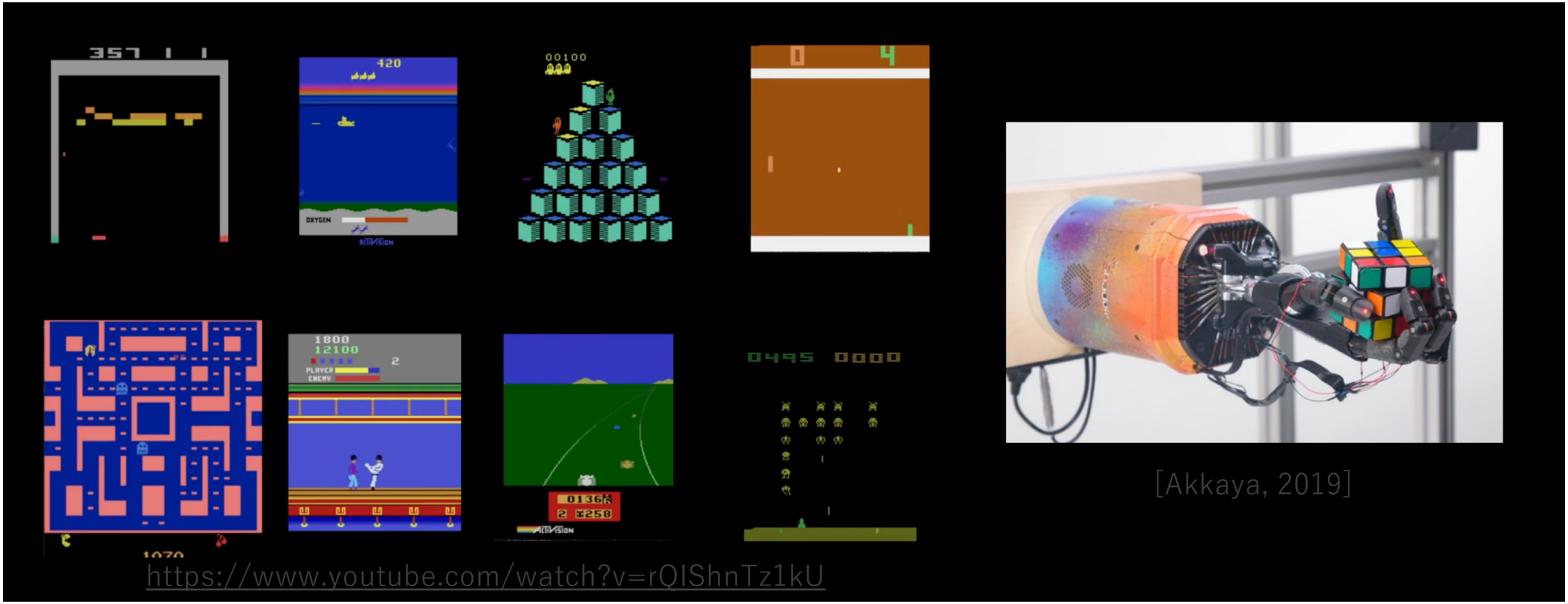
Can Increasing Input Dimensionality Improve Deep Reinforcement Learning?

Kei Ota[1], Tomoaki Oiki[1], Devesh K. Jha[2],
Toshisada Mariyama[1], and Daniel Nikovski[2]

1. Mitsubishi Electrics, Kanagawa, JP
2. Mitsubishi Electric Research Labs, MA, US.

Introduction

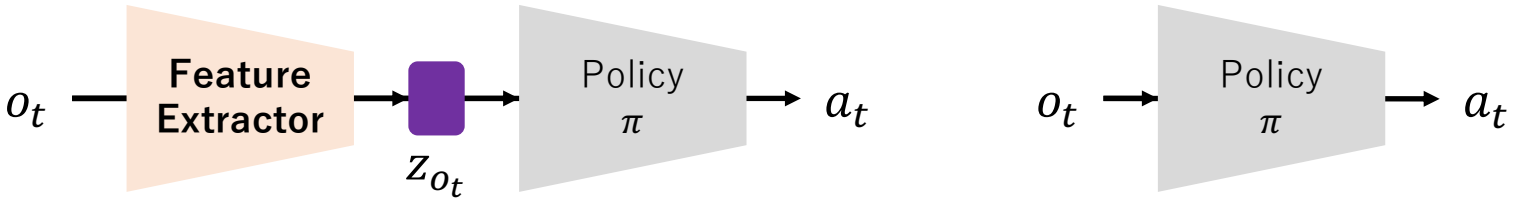
- Deep RL algorithms have achieved impressive success
  - ✓ Can solve complex tasks
  - X Learning representations requires a large amount of data



[Akkaya, 2019]

Introduction

- Deep RL algorithms have achieved impressive success
  - ✓ Can solve complex tasks
  - X Learning representations requires a large amount of data
- State Representation Learning (SRL)
  - Learned features are in low dimension, evolve through time, and are influenced by actions of an agent
  - The lower the dimensionality, the faster and better RL algorithms will learn



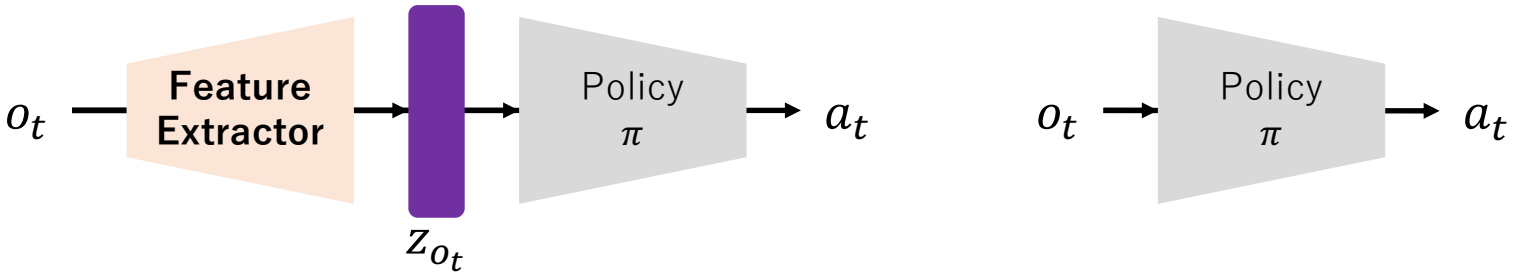
SRL + RL

Standard RL

MITSUBISHI ELECTRIC
Changes for the Better

Introduction

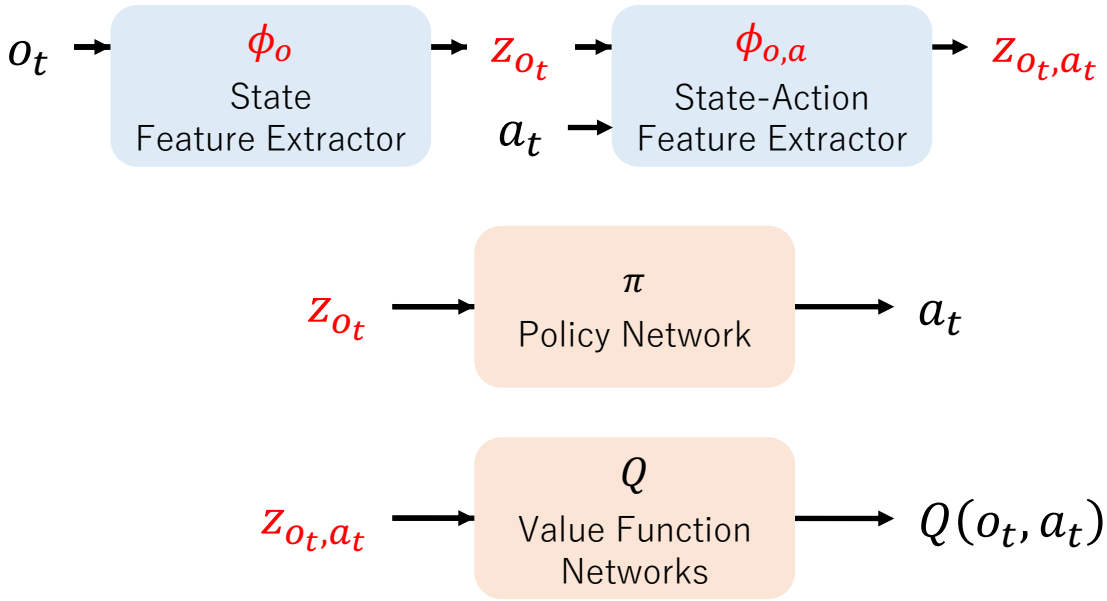
- Deep RL algorithms have achieved impressive success
  - ✓ Can solve complex tasks
  - X Learning representations requires a large amount of data
- State Representation Learning (SRL)
  - Learned features are in low dimension, evolve through time, and are influenced by actions of an agent
  - The lower the dimensionality, the faster and better RL algorithms will learn



Can Increasing Input Dimensionality Improve Deep RL?

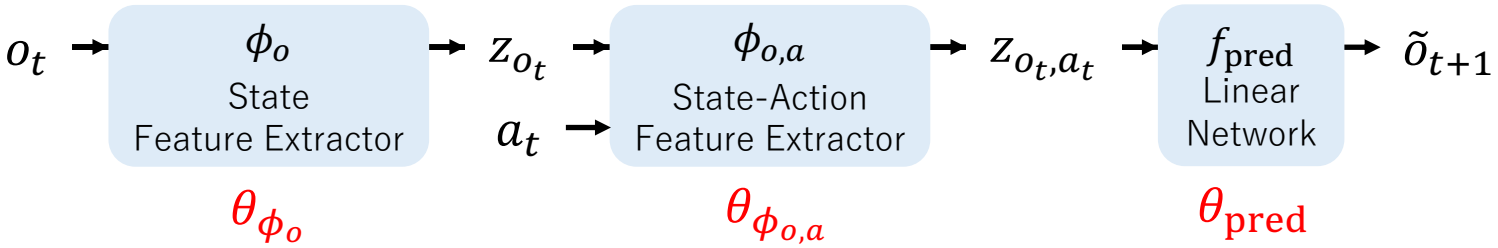
OFENet: Online Feature Extractor Network

- OFENet
  - Train feature extractor network ϕ_o and $\phi_{o,a}$ that produces **high-dimensional** representation z_{o_t} and z_{o_t,a_t}



OFENet: Online Feature Extractor Network

- OFENet
  - Train feature extractor network ϕ_o and $\phi_{o,a}$ that produces **high-dimensional** representation z_{o_t} and z_{o_t,a_t}



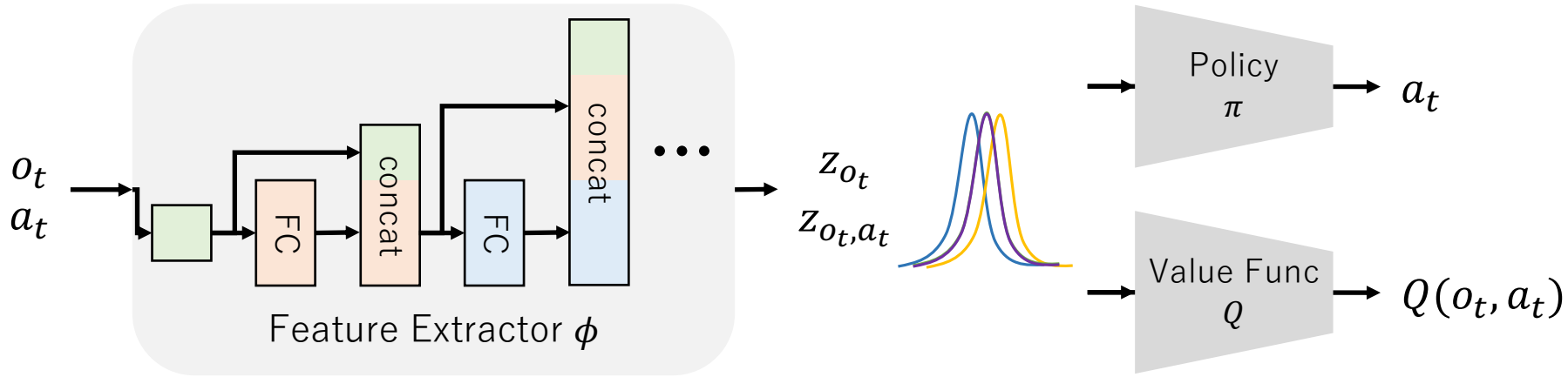
  - Optimize $\theta_{\text{aux}} = \{\theta_{\phi_o}, \theta_{\phi_{o,a}}, \theta_{\text{pred}}\}$ by learning to predict next state

$$L_{\text{aux}} = \mathbb{E}_{(o_t,a_t)\sim p,\pi}\left[\left\| f_{\text{pred}}\left(z_{o_t,a_t}\right) - o_{t+1}\right\|^2\right]$$

  - Increasing the search space allows the agent to learn much more complex policies

Network Architecture

- What is best architecture to extract features?
  - Deeper networks: optimization ability and expressiveness
  - Shallow layers: physically meaningful output
- MLP DenseNet
  - Combine advantages of deep layers and shallow layers



  - Use **Batch Normalization** to suppress changes in input distributions

MITSUBISHI ELECTRIC
Changes for the Better

Experiments

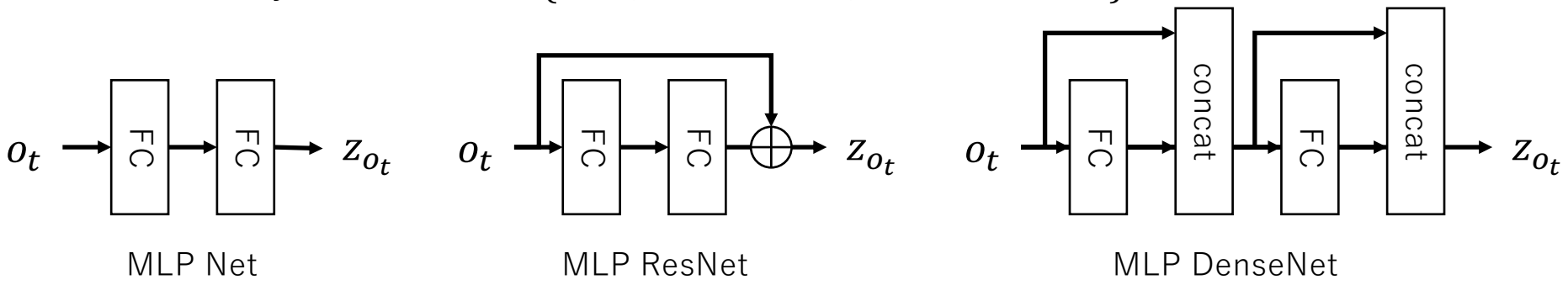
1. **What is a good architecture** that learns effective state and state-action representations for training better RL agents?

2. **Can OFENet learn more sample efficient and better performant** polices when compared to some of the state-of-the-art techniques?

3. **What leads to the performance gain** obtained by OFENet?

What is a good architecture?

- Compare aux. score and actual RL score to search a good architecture from:
  - Connectivity architecture: $\{\text{MLP}, \text{MLP ResNet}, \text{MLP DenseNet}\}$



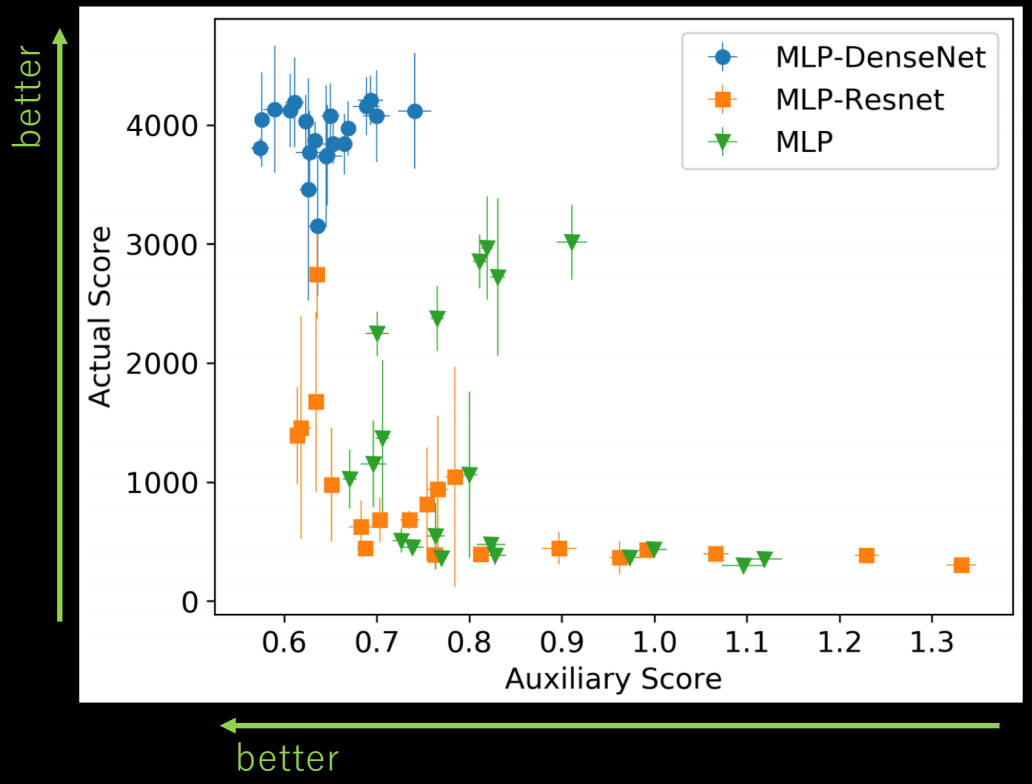
  - Number of layers: $n_{\text{layers}} \in \{1,2,3,4\}$ for MLP, $n_{\text{layers}} \in \{2,4,6,8\}$ for others
  - Activation function: $\{\text{ReLU}, \tanh, \text{Leaky ReLU}, \text{swish}, \text{SELU}\}$

- Aux. score: randomly collect 100K transitions for training, 20K for evaluation

$$L_{aux} = \mathbb{E}_{(o_t,a_t)\sim p,\pi}\left[\left\| f_{pred}\left(z_{o_t,a_t}\right) - o_{t+1}\right\|^2\right]$$

- Actual score: measure returns of SAC agent with 500K steps training

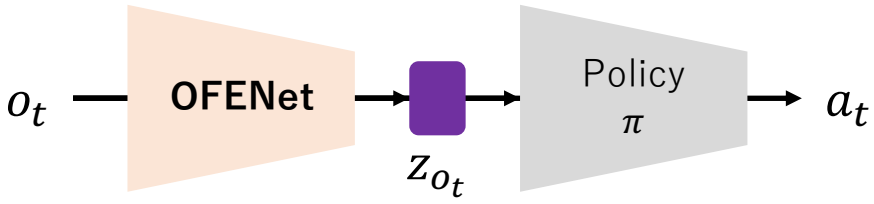
What is a good architecture?



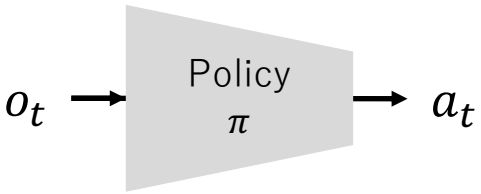
- **MLP-DenseNet** consistently achieves higher actual score
- Smaller the aux. score, better the actual score
- We can select architecture with the smallest aux. score without solving heavy RL problem!

More sample efficient and better performant polices?

- Measure performance of SAC, TD3, and PPO with and without OFENet
  - No changes in hyperparameters for each algorithm

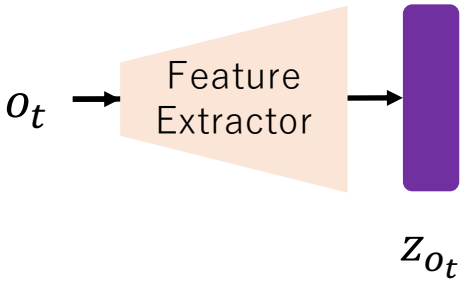


OFENet representation

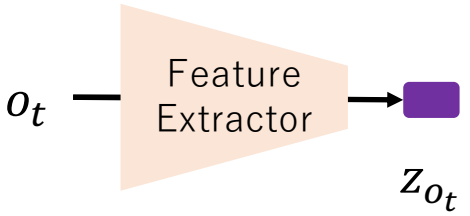


Raw observation

- Compare to closest work: ML-DDPG [Munk2016]
  - **Reduce the dimension** of the observation to one third of its original

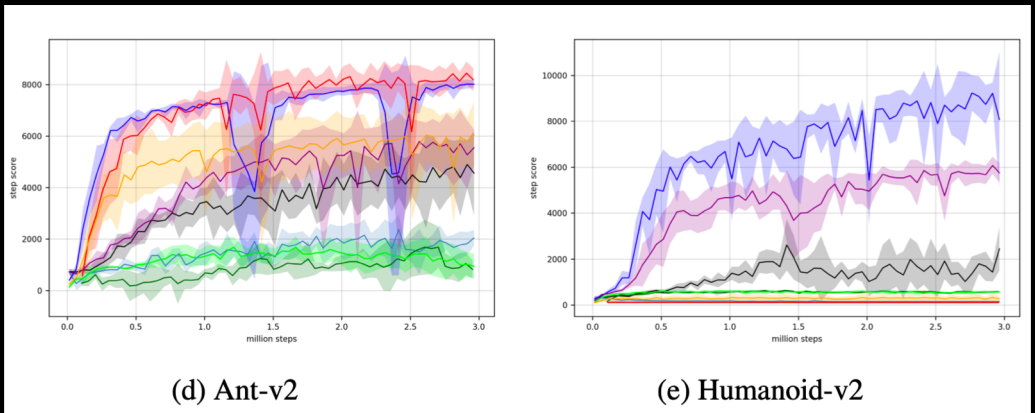


OFENet



ML-DDPG

More sample efficient and better performant polices?



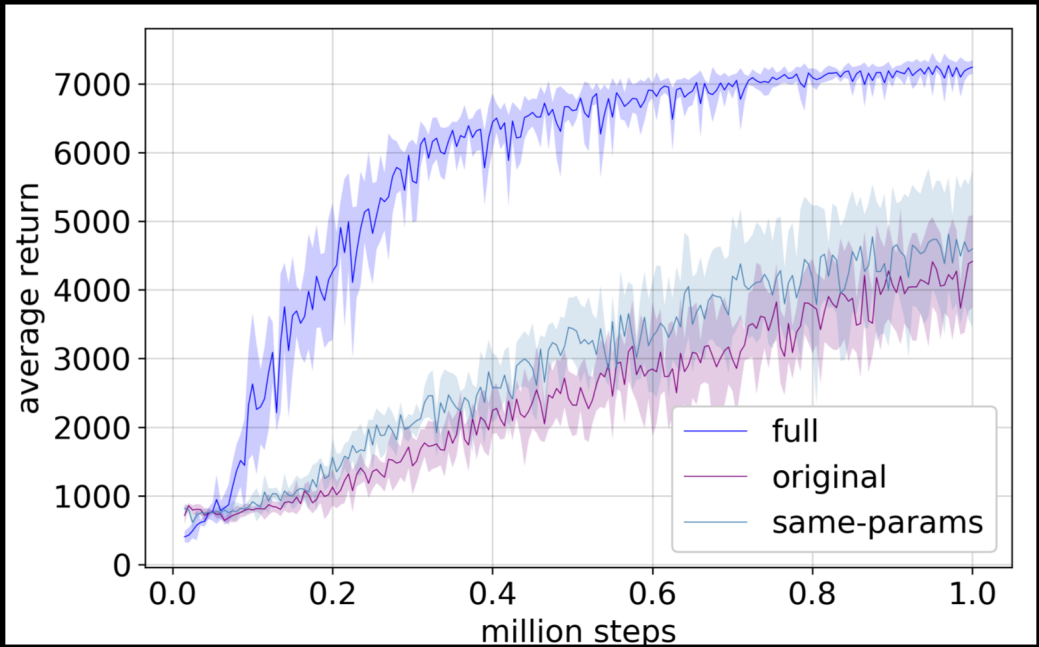
(d) Ant-v2 (e) Humanoid-v2

- OFENet improves sample efficiency and returns without changing any hyperparameters
- OFENet effectively learns meaningful features

	SAC				TD3		PPO	
	OFE (OURS)	Original	ML-SAC (1/3)	ML-SAC (OFE like)	OFE (OURS)	Original	OFE (OURS)	Original
HOPPER-V2	**3511.6**	3316.6	750.5	868.7	3488.3	**3613.0**	**2525.6**	1753.5
WALKER2D-V2	**5237.0**	3401.5	667.4	627.4	**4915.1**	4515.6	**3072.1**	3016.7
HALFCHEETAH-V2	**16964.1**	14116.1	1956.9	11345.5	**16259.5**	13319.9	**3981.8**	2860.4
ANT-V2	**8086.2**	5953.1	4950.9	2368.3	**8472.4**	6148.6	**1782.3**	1678.9
HUMANOID-V2	**9560.5**	6092.6	3458.2	331.7	120.6	**345.2**	**670.3**	652.4

MITSUBISHI ELECTRIC
Changes for the Better

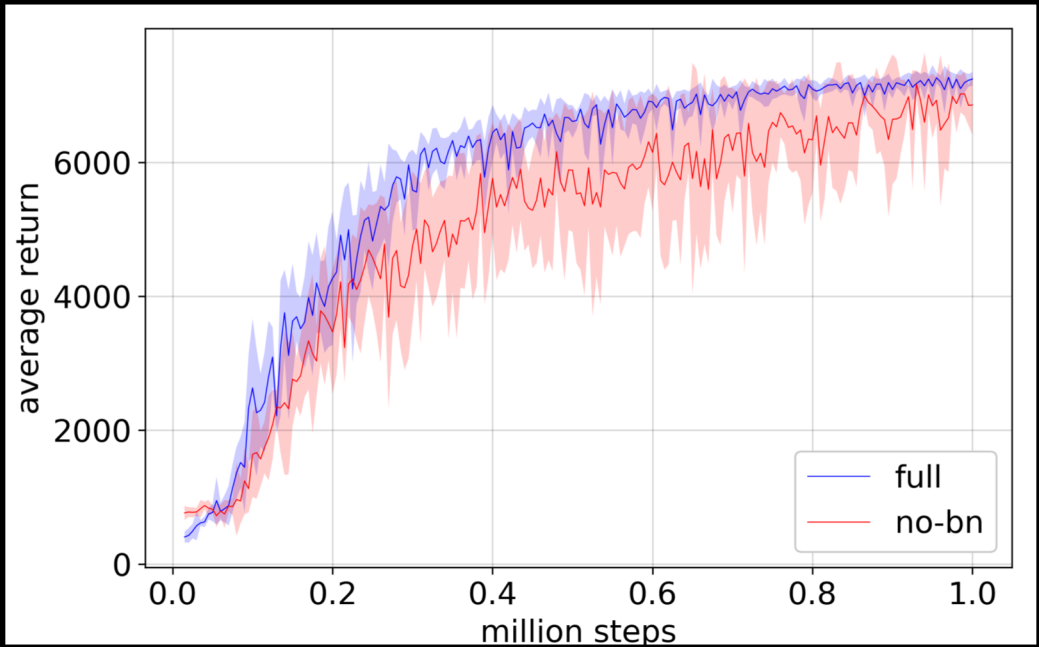
What leads to the performance gain?



- Just increasing network size doesn't improve performance

MITSUBISHI ELECTRIC
Changes for the Better

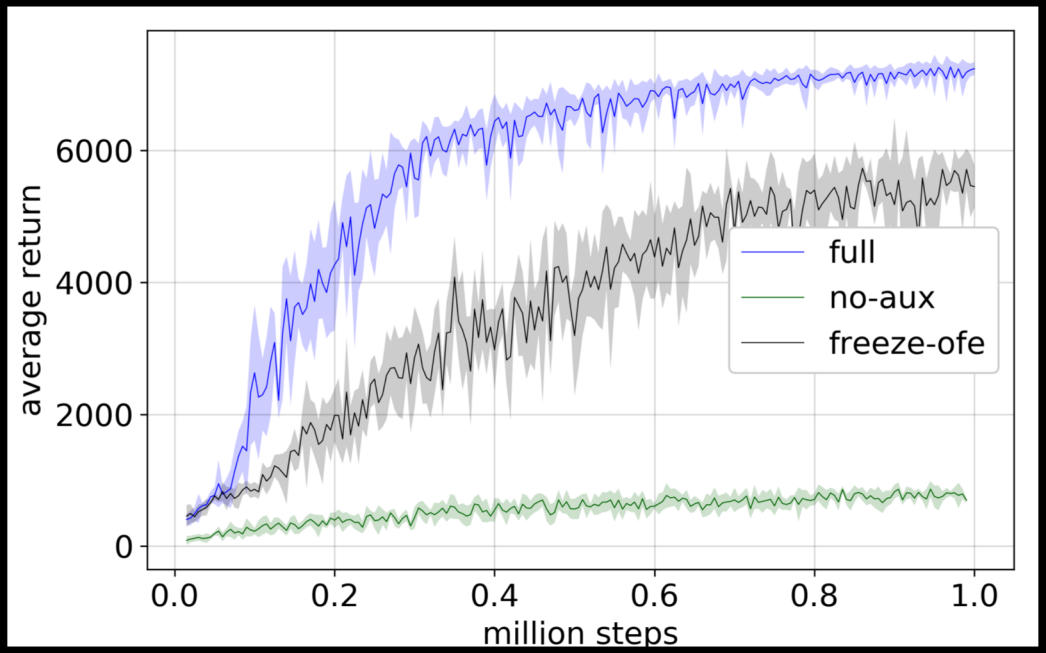
What leads to the performance gain?



- Just increasing network size doesn't improve performance
- **BN stabilizes training**

MITSUBISHI ELECTRIC
Changes for the Better

What leads to the performance gain?



- Just increasing network size doesn't improve performance
- BN stabilizes training
- **Decoupling feature extraction and control policy is important**
- **Online SRL handles unknown distribution during training**

MITSUBISHI ELECTRIC
Changes for the Better

Conclusion

- Proposed Online Feature Extractor Network (OFENet)
  - Provides much higher-dimensional representation
  - Demonstrated OFENet can significantly accelerate RL

- OFENet can be used as New RL tool box
  - Just put OFENet as base layer of RL algorithms
  - No need to tune hyperparameters of original algorithms!
  - Code link: www.merl.com/research/license/OFENet

Can increasing input dimensionality improve deep RL?

Yes, it can!