

テレビゲームを用いた 見まね学習の比較

Comparison of imitation learning
using video games

西田 圭吾 大阪大学
田村 陵大 同志社大学大学院
三浦 拓也 大阪大学
玉城 貴也 同志社大学

模倣学習

- 人の言語獲得,楽器演奏,運動選手の運動フォーム等は後天的に学習によって獲得される技能
- これらの行動は他人の模倣により、効率的に高い技能を獲得することができる
- その行動パターン生成・維持は運動出力とそれをモニターし、他人の行動と比較するフィードバック制御を行うことで実現される

小鳥のさえずり模倣学習

鳴禽と呼ばれる小鳥類は、言語を学習する人間と同様、「さえずり」という複雑な音声パターンを他個体からの模倣により発達させる

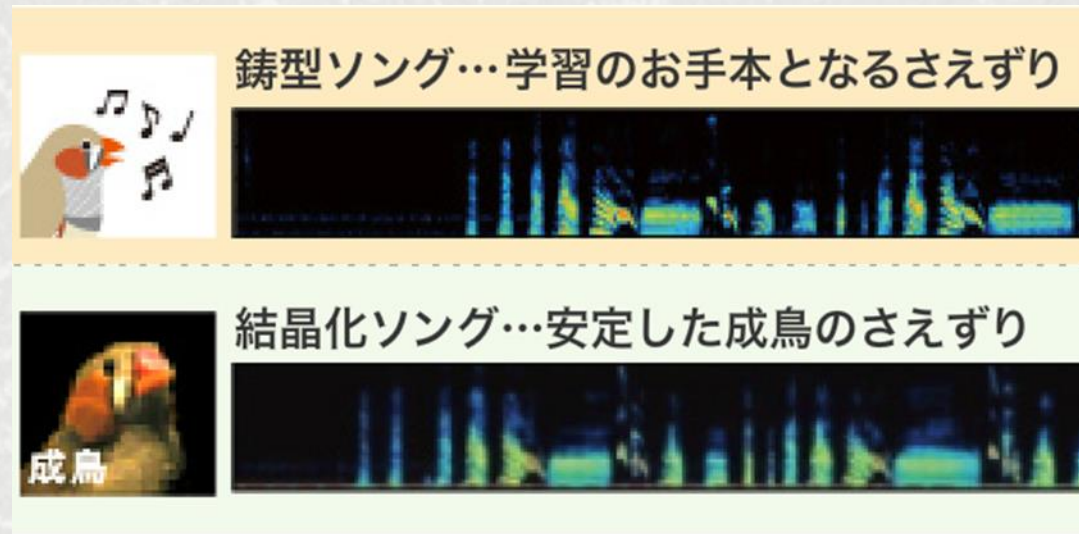
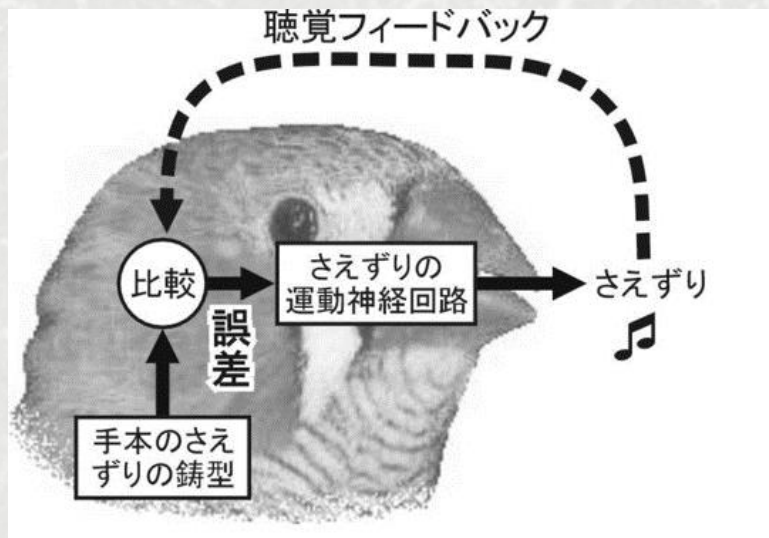
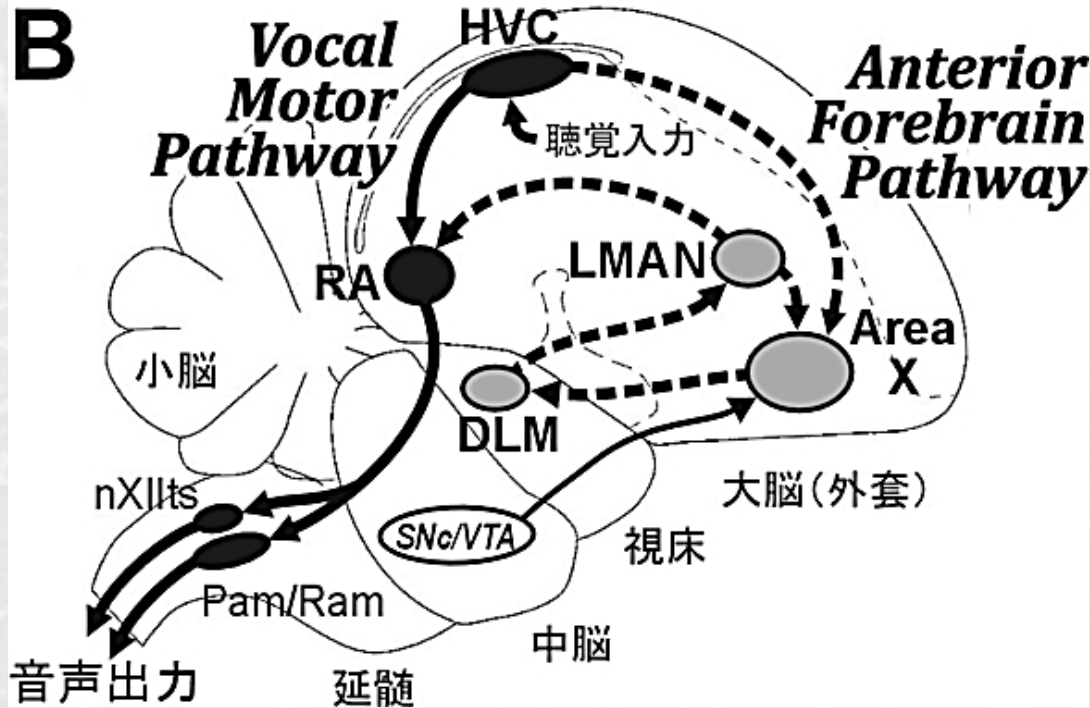


図 一部改変

小鳥のさえずり模倣学習の主な神経回路



Vocal Motor
Pathway(直接制御系)
・運動神経経路に対応

Anterior Forebrain
Pathway(迂回投射系)
・大脳皮質-大脳基底
核ループに対応

各神経核との大まかな対応

HVC : 運動前野

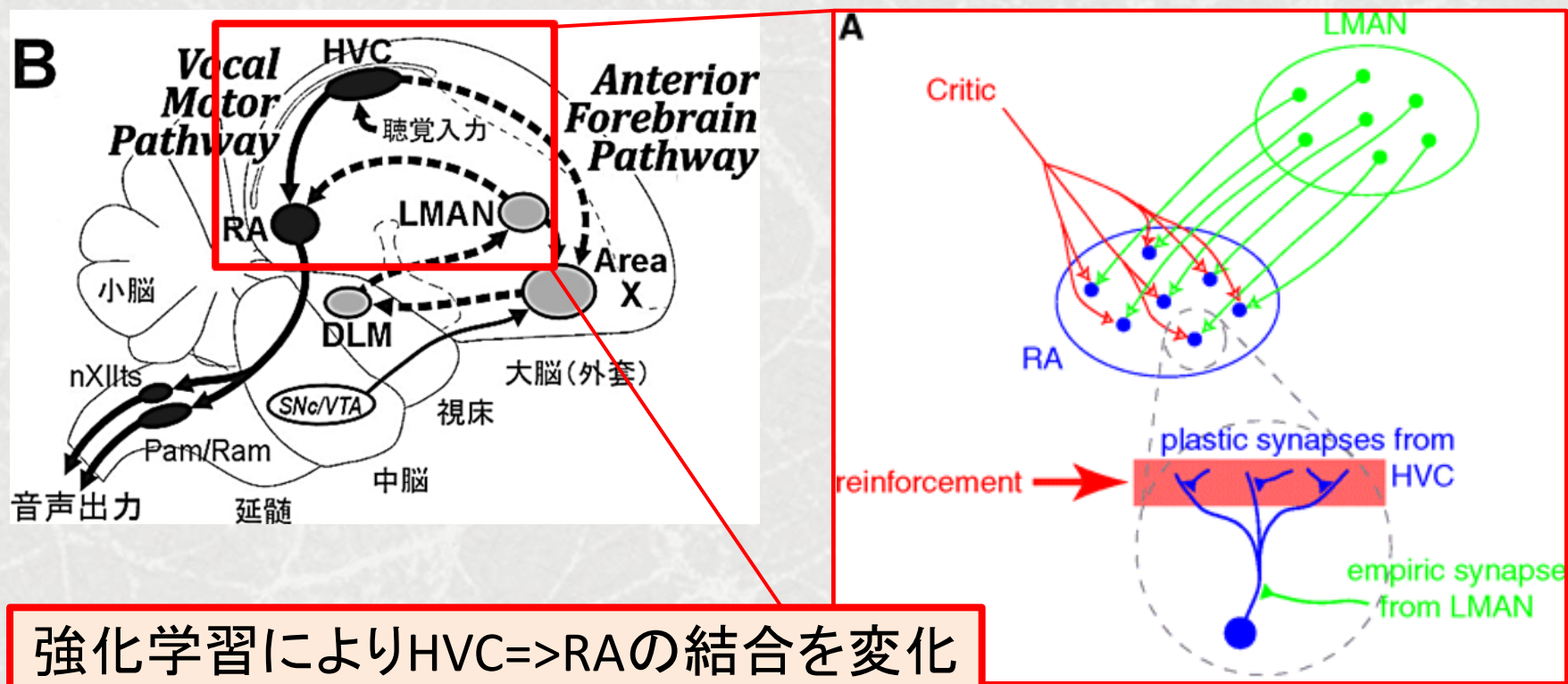
RA : 一次運動野

Area X : 大脳基底核

DLM : 視床

LMAN : 大脳皮質

小鳥のさえずり模倣学習モデル



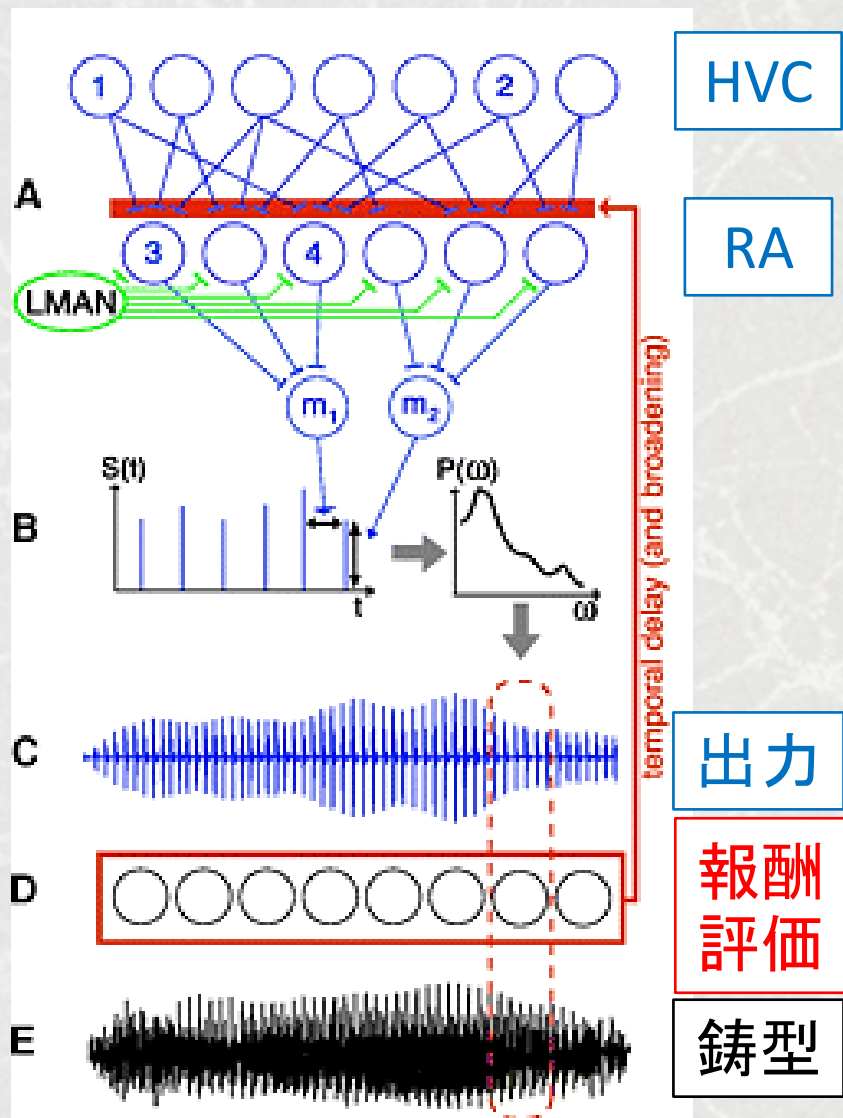
HVC :行動の時系列遷移パターンを出力

RA :モータへの行動出力

LMAN:試行パターンの揺らぎを生成(探索の生成)

Critic :鋳型(お手本)と比較した際に与えられる報酬

鑄型(お手本)との比較による報酬生成



出力と鑄型を比較することで
報酬を決定する

疑問

鑄型は脳内において如何に
表現され学習に寄与するのか?



仮説

脳内で特徴量抽出し終えた状態
のさえずりパターンを基に想起に
よって再現される鑄型を用いて
強化学習を行う

WBAIハッカソンテーマ

仮説

脳内で特徴量抽出し終えた状態のさえずりパターンを基に想起によって再現される鋳型を用いて強化学習を行う

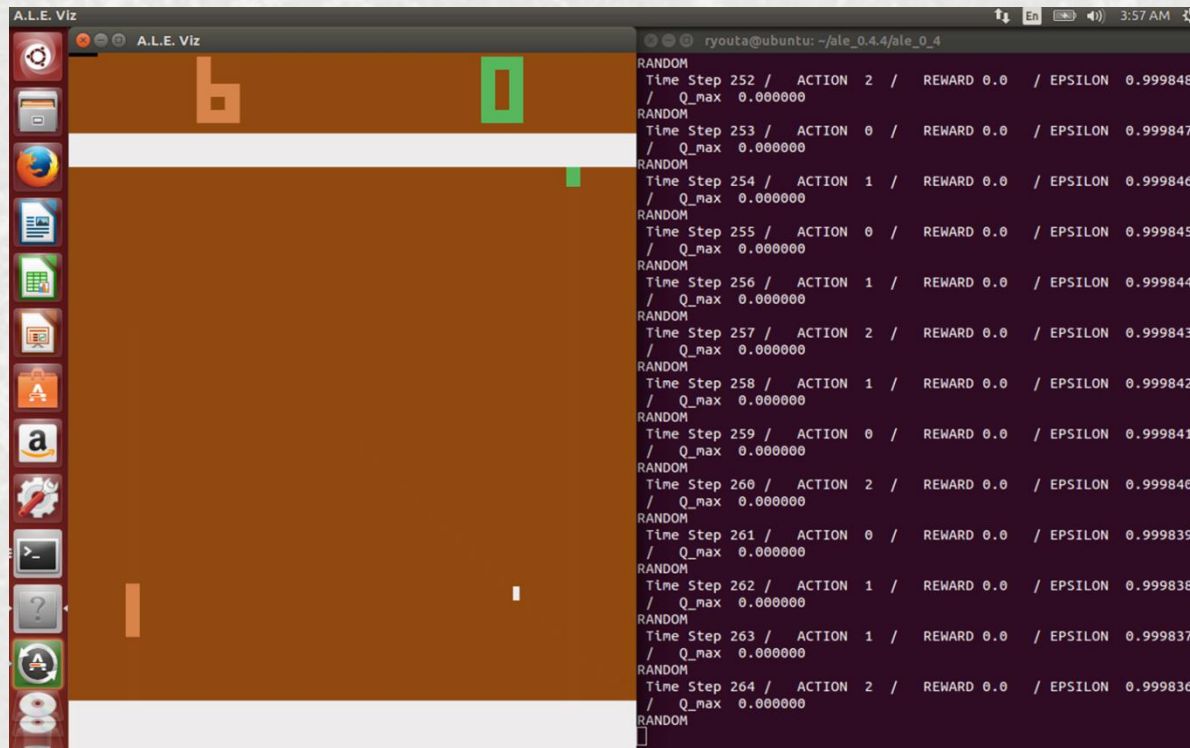


実装

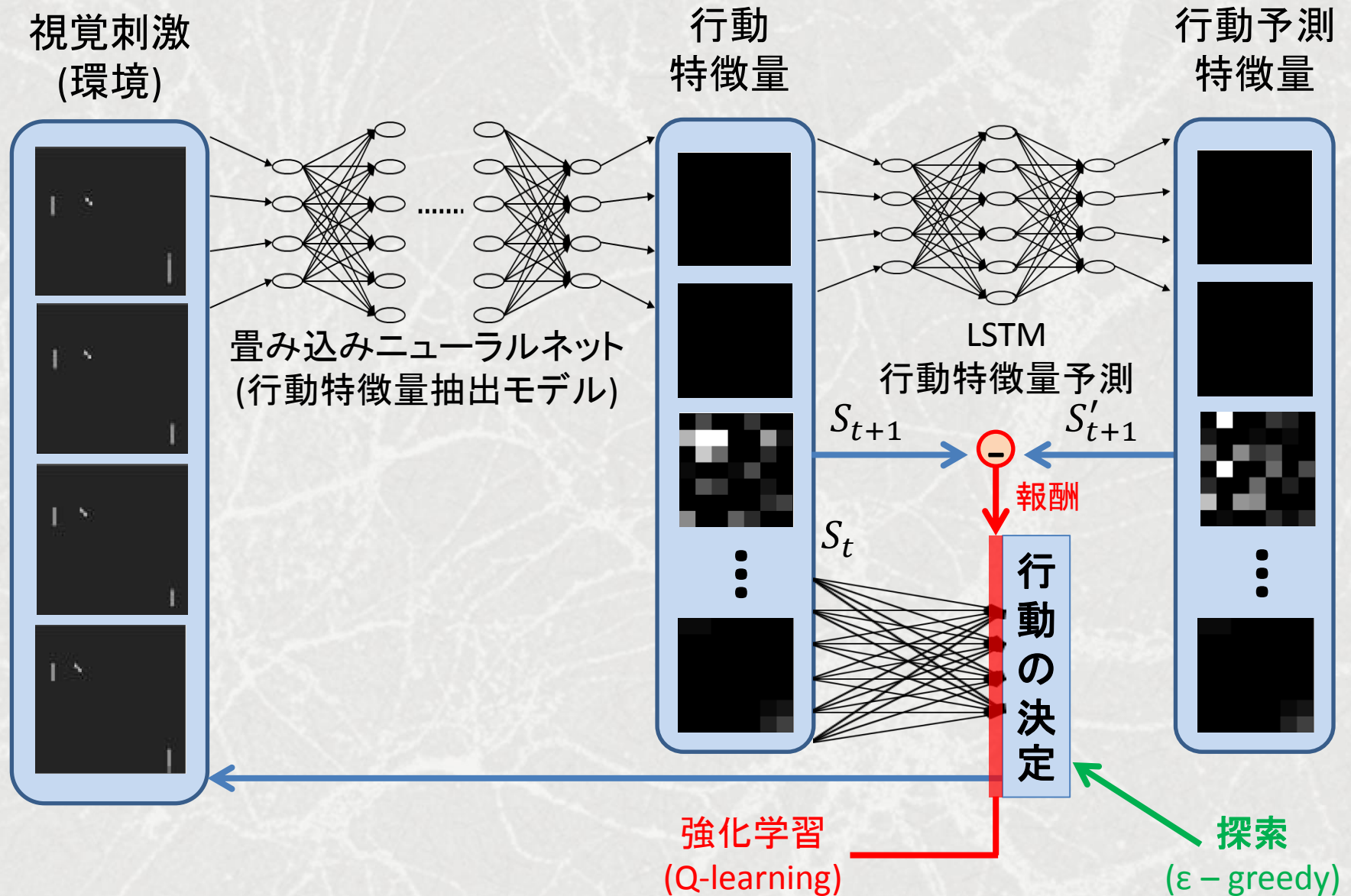
- ①鋳型(お手本)となるエキスパートとその受容野を作製し
- ②LSTMを用いて想起の内部モデルを獲得することで
- ③模倣学習を強化学習により再現できるかを確かめる

Deep Q-Network

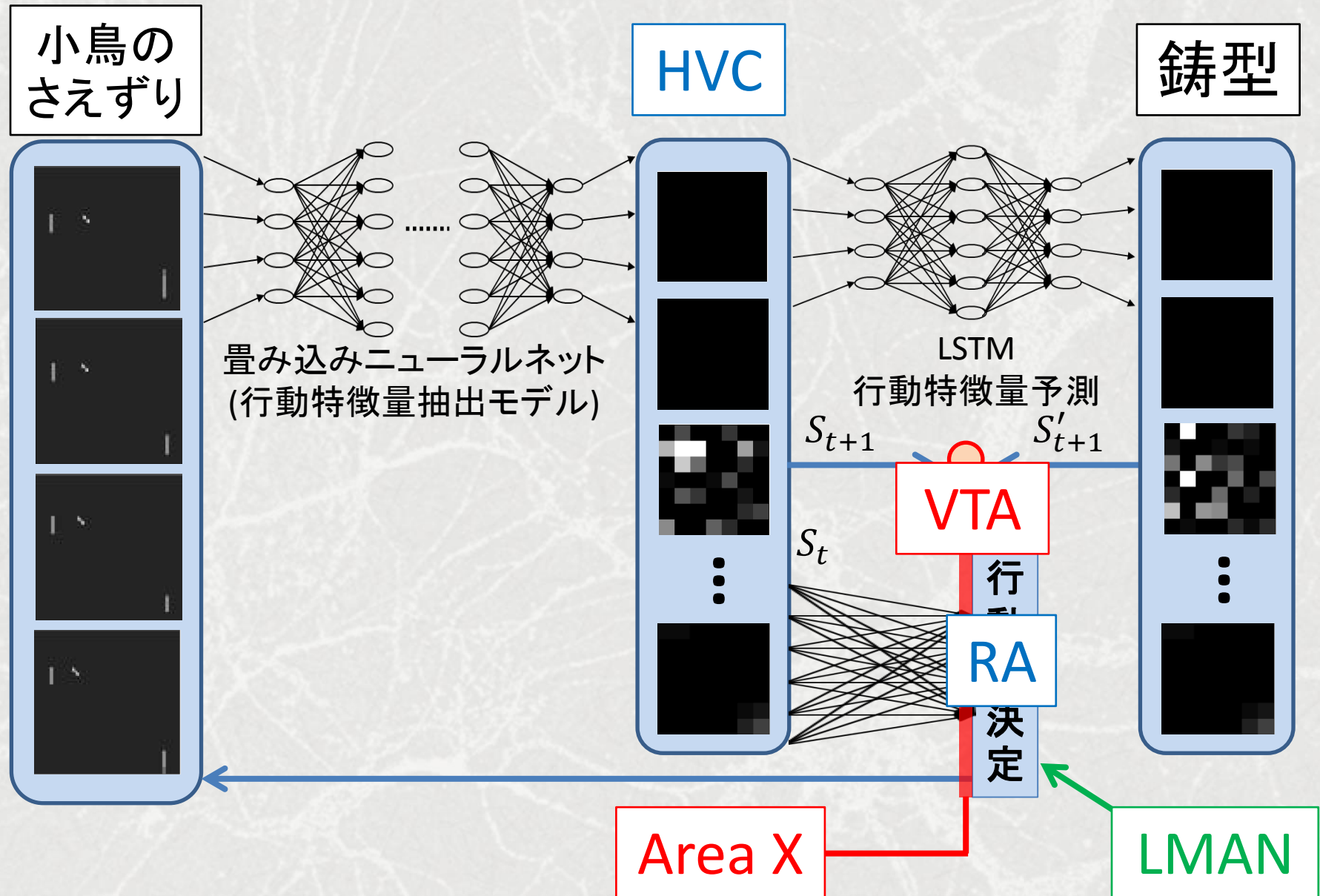
- 『Atari 2600』のゲーム49本を学習させたところ、その半数以上で人間に匹敵、時には上回るスコアを記録した
- これを利用して小鳥の模倣学習モデルを実装



実装したアーキテクチャ

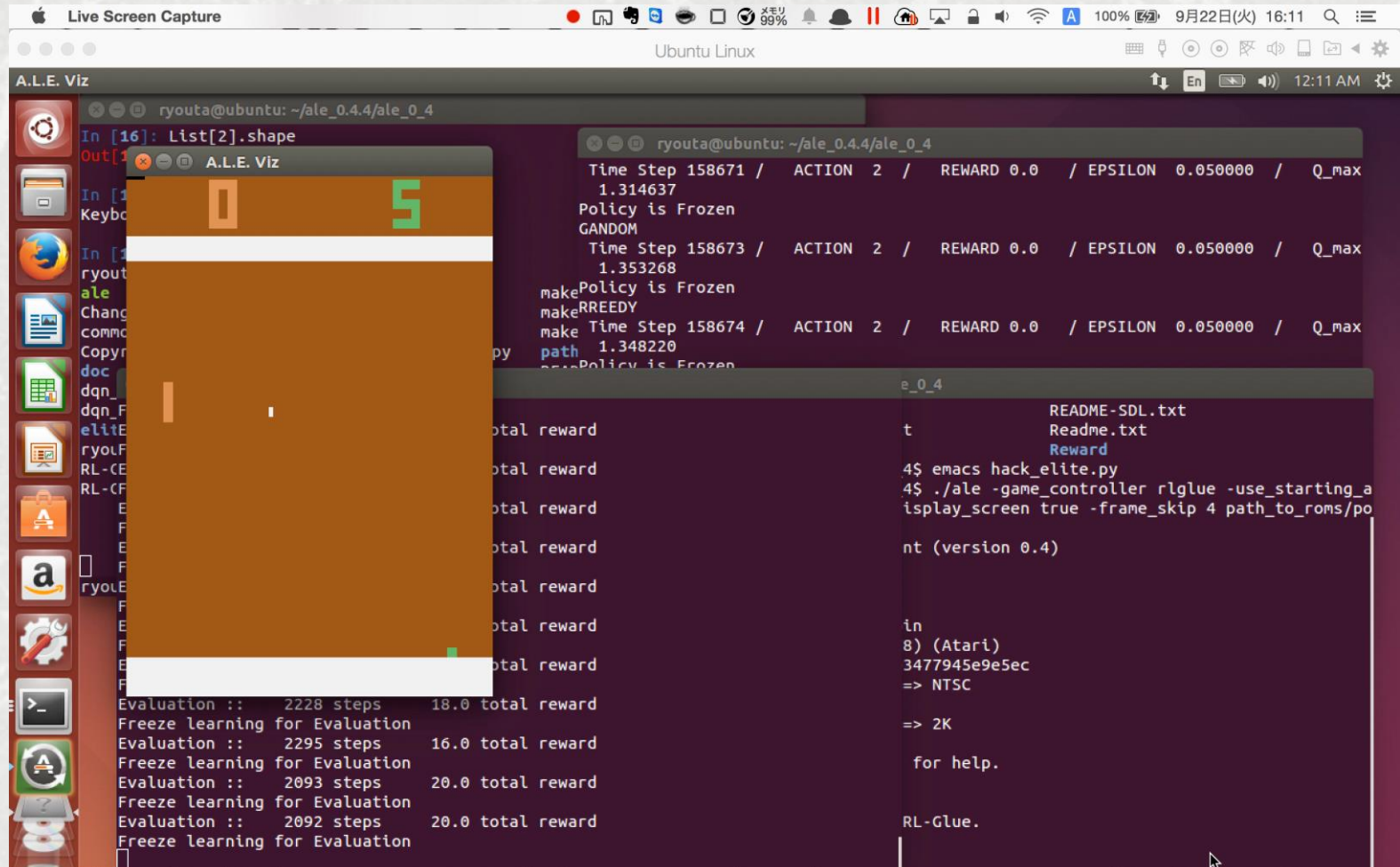


実装したアーキテクチャ(小鳥との対応)



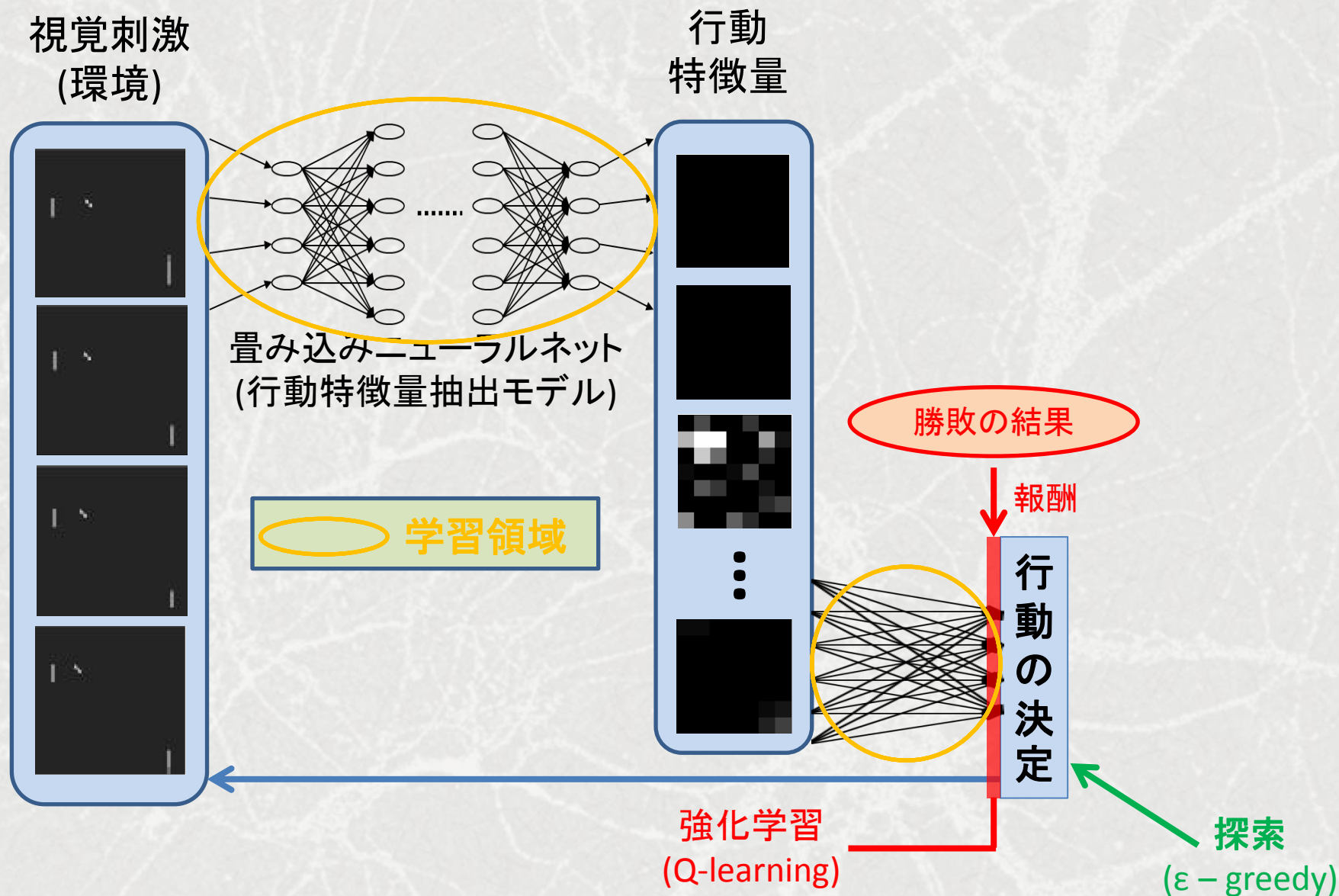
エキスパート(先生)の育成

- Deep Q-Networkを使ったエキスパートの獲得

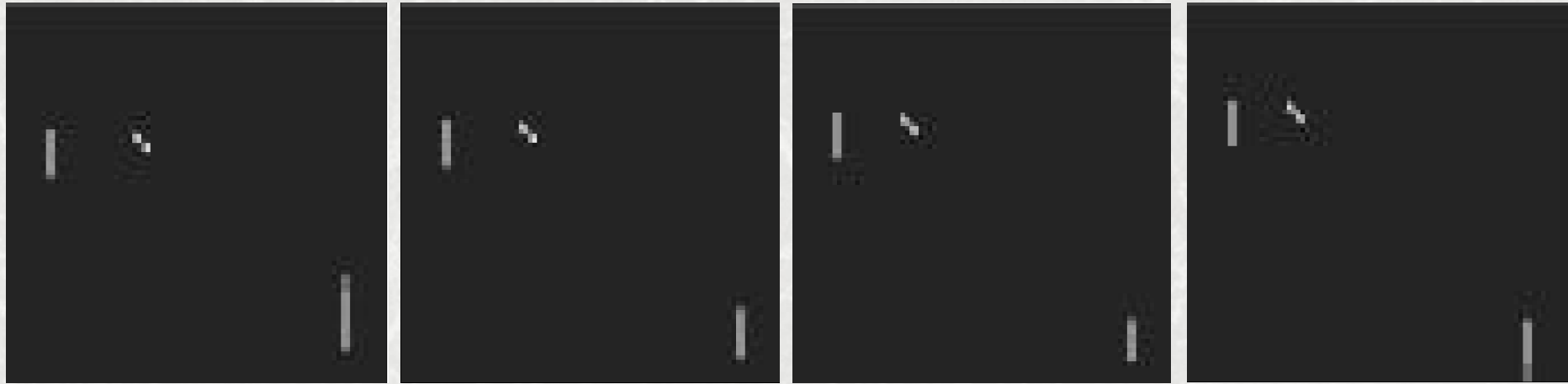


```
A.L.E. Viz
ryouta@ubuntu: ~/ale_0.4.4/ale_0_4
In [16]: List[2].shape
Out[16]: (10, 10)
A.L.E. Viz
ryouta@ubuntu: ~/ale_0.4.4/ale_0_4
Time Step 158671 / ACTION 2 / REWARD 0.0 / EPSILON 0.050000 / Q_max 1.314637
Policy is Frozen
GANDOM
Time Step 158673 / ACTION 2 / REWARD 0.0 / EPSILON 0.050000 / Q_max 1.353268
Policy is Frozen
make REEDY
make Time Step 158674 / ACTION 2 / REWARD 0.0 / EPSILON 0.050000 / Q_max 1.348220
Policy is Frozen
e_0_4
total reward
t
Readme.txt
Readme.txt
Reward
4$ emacs hack_elite.py
4$ ./ale -game_controller rlg glue -use_starting_a
isplay_screen true -frame_skip 4 path_to_roms/po
nt (version 0.4)
in
8) (Atari)
3477945e9e5ec
=> NTSC
=> 2K
for help.
RL-Glue.
```

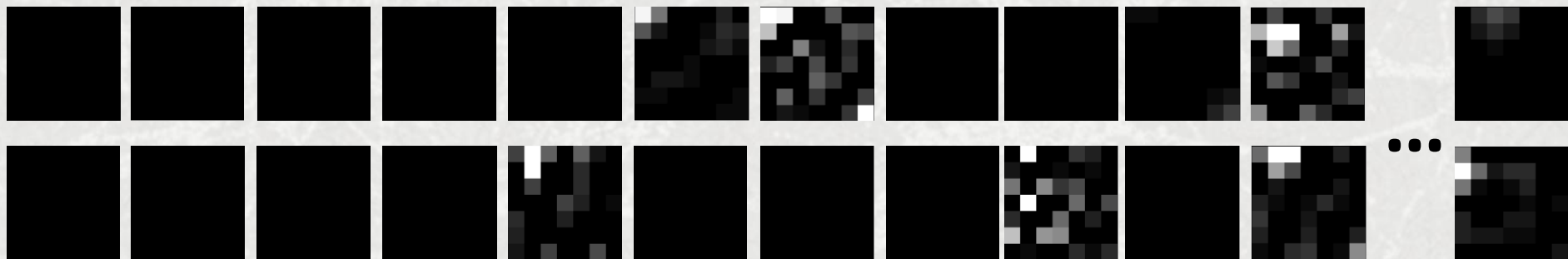

① エキスパートと受容野の作製(DQNの実行)



エキスパートの特徴量空間

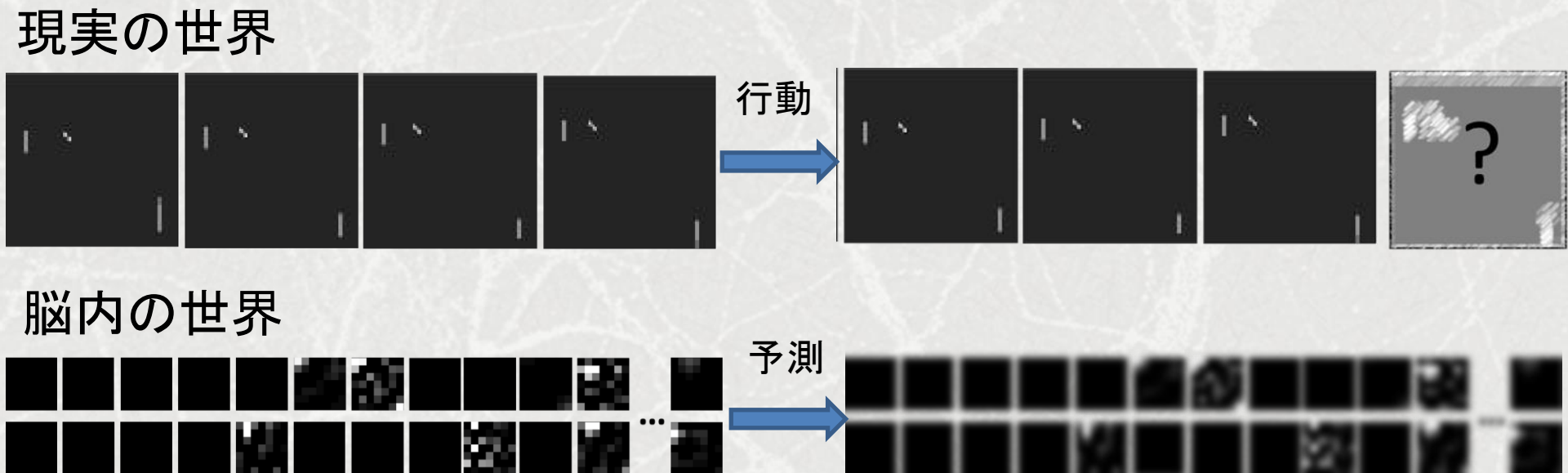


CNNによる特徴量抽出



7 pix × 7 pix × 64 個の出力

先生の行動予測学習



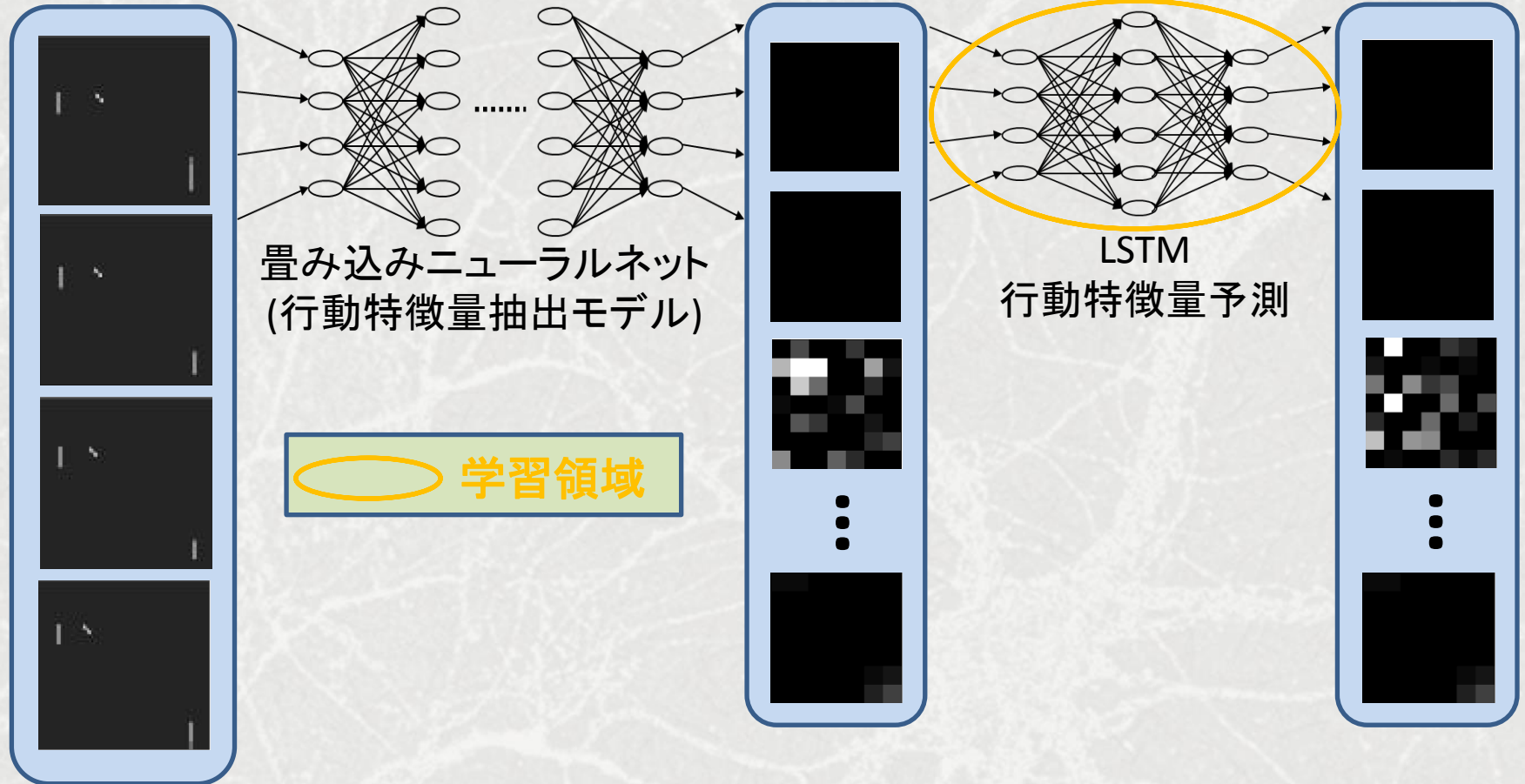
学習は試合の決着がつくタイミングで区切って
LSTMの学習を行った

LSTMを用いた想起の内部モデル獲得

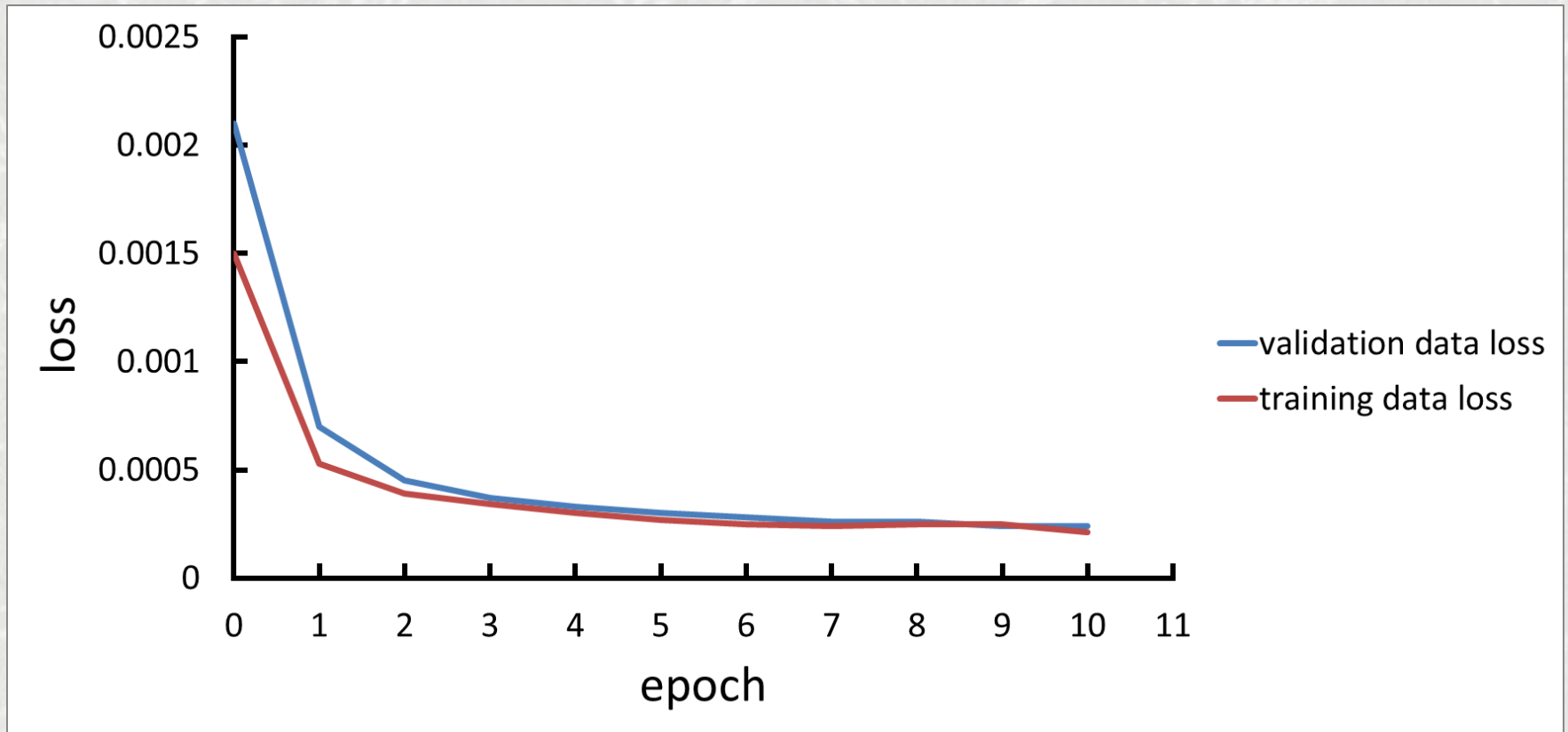
エキスパート
プレイ動画

行動
特徴量

行動予測
特徴量

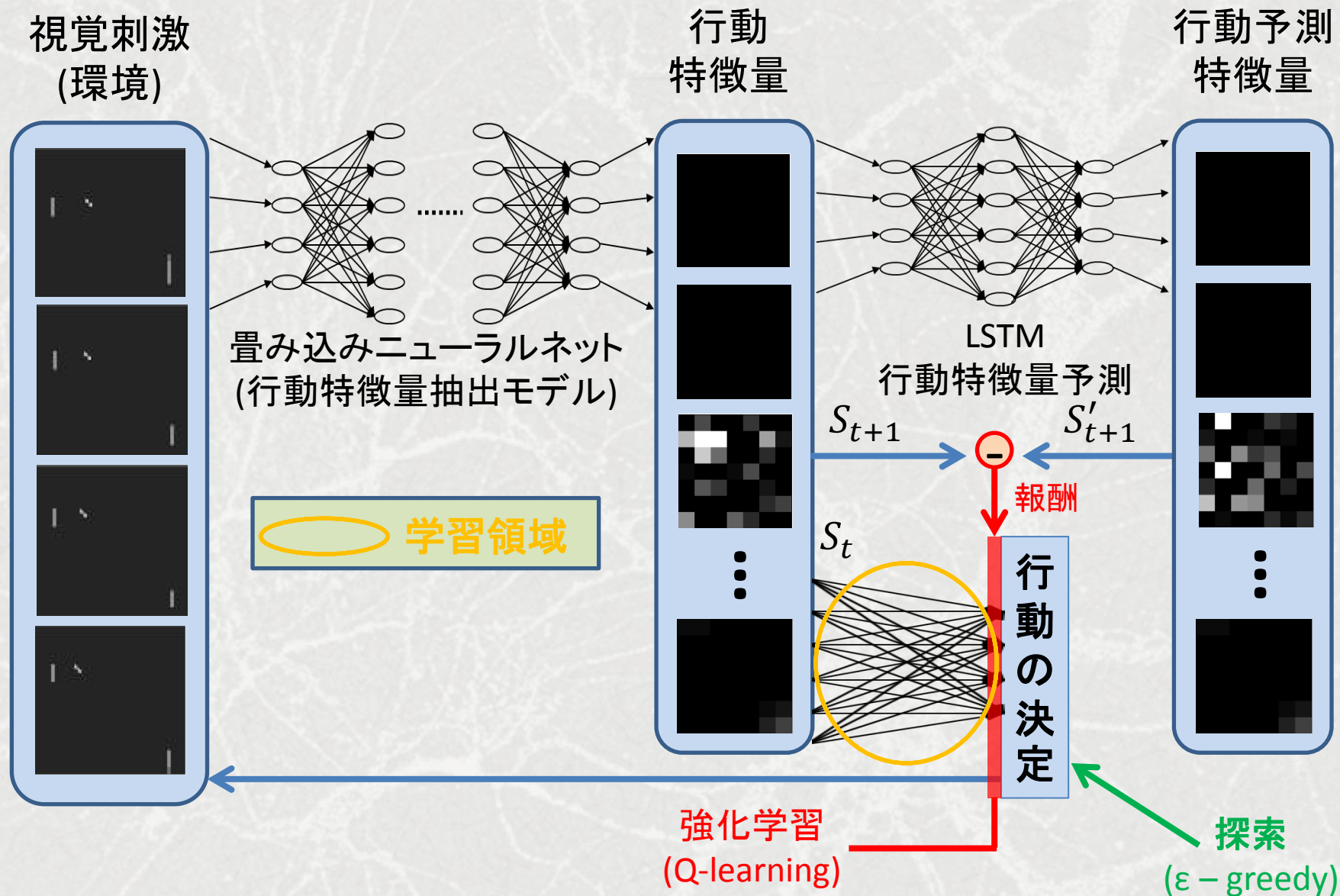


用いたLSTM単体の性能評価

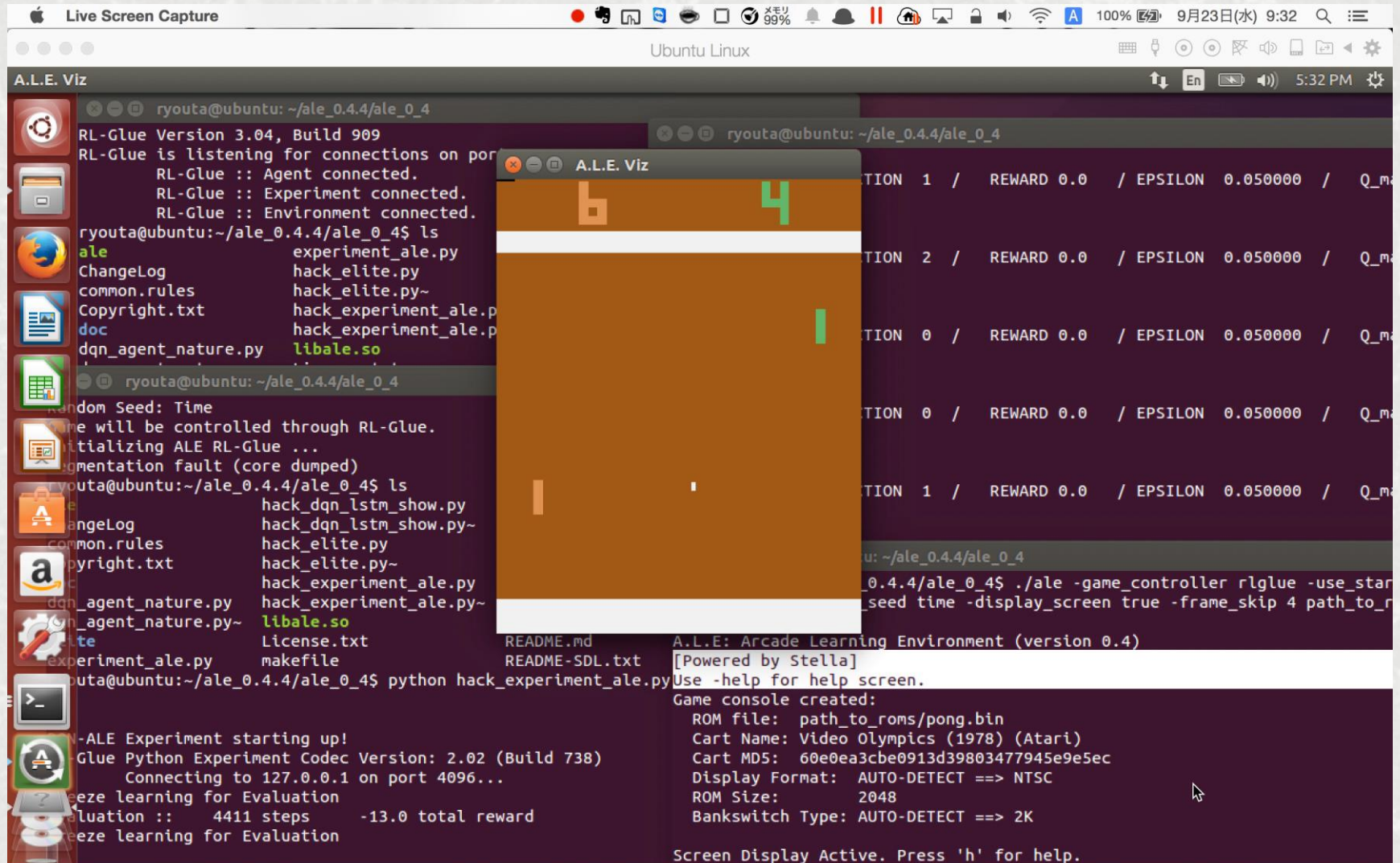


パラメータの不足により、十分な表現力が無い可能性があったが続行した

模倣学習を強化学習により再現



DQN+LSTM(1 epoch) 3時間の学習

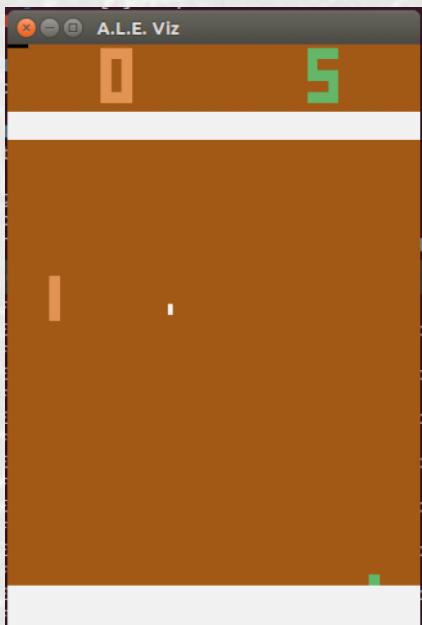


```
Live Screen Capture
Ubuntu Linux
A.L.E. Viz
ryouta@ubuntu: ~/ale_0.4.4/ale_0_4
RL-Glue Version 3.04, Build 909
RL-Glue is listening for connections on port 4096.
RL-Glue :: Agent connected.
RL-Glue :: Experiment connected.
RL-Glue :: Environment connected.
ryouta@ubuntu:~/ale_0.4.4/ale_0_4$ ls
ale               experiment_ale.py
ChangeLog         hack_elite.py
common.rules      hack_elite.py~
Copyright.txt     hack_experiment_ale.py
doc               hack_experiment_ale.py~
dqn_agent_nature.py libale.so
ryouta@ubuntu:~/ale_0.4.4/ale_0_4$
Random Seed: Time
The game will be controlled through RL-Glue.
Initializing ALE RL-Glue ...
Segmentation fault (core dumped)
ryouta@ubuntu:~/ale_0.4.4/ale_0_4$ ls
ale               experiment_ale.py
ChangeLog         hack_dqn_lstm_show.py
common.rules      hack_dqn_lstm_show.py~
Copyright.txt     hack_elite.py
dqn_agent_nature.py hack_elite.py~
dqn_agent_nature.py~ hack_experiment_ale.py
hack_experiment_ale.py~ libale.so
License.txt       makefile
ryouta@ubuntu:~/ale_0.4.4/ale_0_4$ python hack_experiment_ale.py
N-ALE Experiment starting up!
RL-Glue Python Experiment Codec Version: 2.02 (Build 738)
Connecting to 127.0.0.1 on port 4096...
Freeze learning for Evaluation
Duration :: 4411 steps    -13.0 total reward
Freeze learning for Evaluation

A.L.E. Viz
6 4
A.L.E. Arcade Learning Environment (version 0.4)
[Powered by Stella]
Use -help for help screen.
Game console created:
ROM file: path_to_roms/pong.bin
Cart Name: Video Olympics (1978) (Atari)
Cart MD5: 60e0ea3cbe0913d39803477945e9e5ec
Display Format: AUTO-DETECT ==> NTSC
ROM Size: 2048
Bankswitch Type: AUTO-DETECT ==> 2K
Screen Display Active. Press 'h' for help.
```

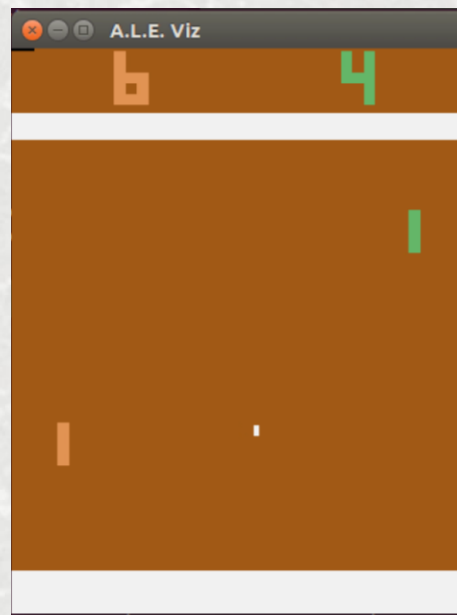
ハッカソンで得られたプレイ動画

エキスパート
(先生)



学習時間 12 時間

DQN+LSTM_1epoch
先生を軽くイメージ



学習時間 3 時間

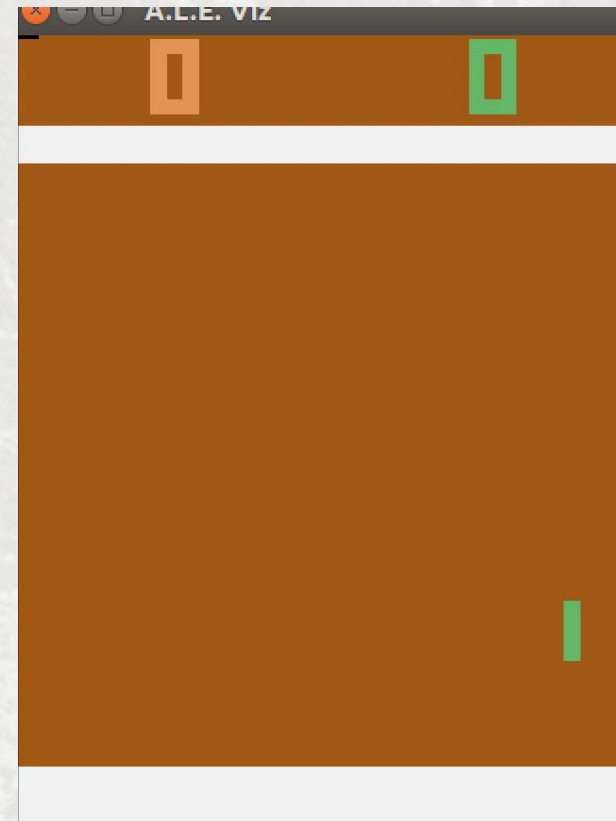
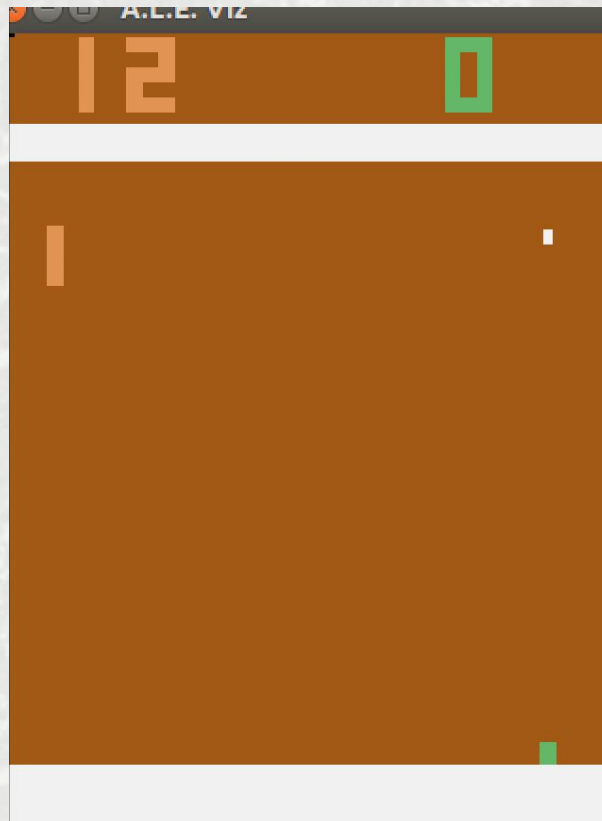
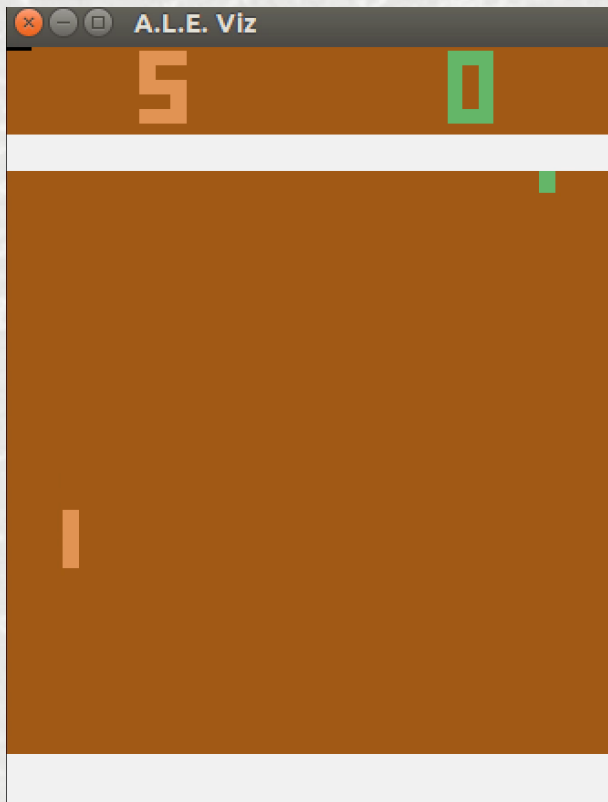
試合に勝利した際の報酬の割合が少なかったために、
球は追うが最後の詰めが甘いエージェントの挙動が確認できる

30分での学習の比較

DQNのみ
完全に模索

DQN+LSTM_1epoch
先生を軽くイメージ

DQN+LSTM_10 epoch
先生を何度もイメージ



DQN+LSTMのエージェントは少ない学習時間である程度、球を追うことができるようになった

結論

- 特徴量データから獲得した内部モデル(鋳型)を用いて、模倣学習を強化学習により実行できることが確認できた
- 内部モデル構築の学習回数を変更したが大きな差は見られなかった(LSTMの過学習?)
- 提案されている小鳥のさえずり模倣学習モデルとは異なる部分(LMANの簡略化)もあり、今後はこのアーキテクチャをたたき台として新たな開発を進めたい

参考文献等

- V. Mnih et al., “Human-level control through deep reinforcement learning”
- Doya K et al., “A computational model of birdsong learning by auditory experience and auditory feedback”
- Ila R. Fiete et al., “Model of Birdsong Learning Based on Gradient Estimation by Dynamic Perturbation of Neural Conductances”
- 小島 哲 “小鳥のさえずり学習の神経機構：大脳基底核経路と強化学習モデル”
- 和多 和宏 “小鳥がさえずるとき脳内では何が起きている？”
- DQNの生い立ち + Deep Q-NetworkをChainerで書いた
<http://qiita.com/Ugo-ama/items/08c6a5f6a571335972d5>

小鳥のさえずり模倣学習モデルとの対応

	小鳥のさえずり模倣学習モデル	ハッカソン実装モデル
エキスパート	他の成鳥のさえずり	DQNで人を超えた実力のプレイ動画
特徴量抽出後の出力	HVC出力	CNN出力
探索	LMAN出力による影響	ϵ - greedy
行動の決定	RA出力	Q値(NNでの近似)による決定
与えられる報酬評価(Critic)	VTA? Area X? (評価方法不明)	LSTM出力とCNN出力との二乗差による評価
鋳型モデル	実空間での周波数スペクトル	LSTMによる想起出力