研究紹介

川本・計良研究室
M2 髙野剛志

CHIBA
UNIVERSITY

Computer Vision
and Machine Learning
Kawamoto Lab.

研究の方向性

研究テーマ (現在)

Atariゲームに対するTransformerベース強化学習のロバスト性検証

実験設定

Decision Transformer × Common Corruptions (Noise)

最終目標

- ❏ Common Corruptionにおける各種ノイズ評価テストにて脆弱性を示す
- ❏ データ拡張訓練により，ロバスト性を向上させる

研究の方向性

研究テーマ (現在)

Atariゲームに対するTransformerベース強化学習のロバスト性検証

実験設定

Decision Transformer × Common Corruptions (Noise)

最終目標

- ❏ Common Corruptionにおける各種ノイズ評価テストにて脆弱性を示す
- ❏ データ拡張訓練により，ロバスト性を向上させる

基礎知識 | Transformer

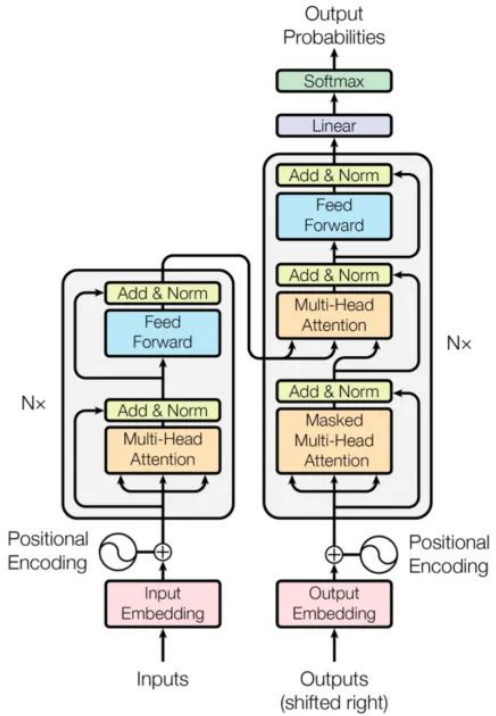


Figure 1: The Transformer - model architecture.

"Attention Is All You Need",2017

Attention is all you need

A Vaswani, N Shazeer, N Parmar… - Advances in neural …, 2017 -
proceedings.neurips.cc Paperpile

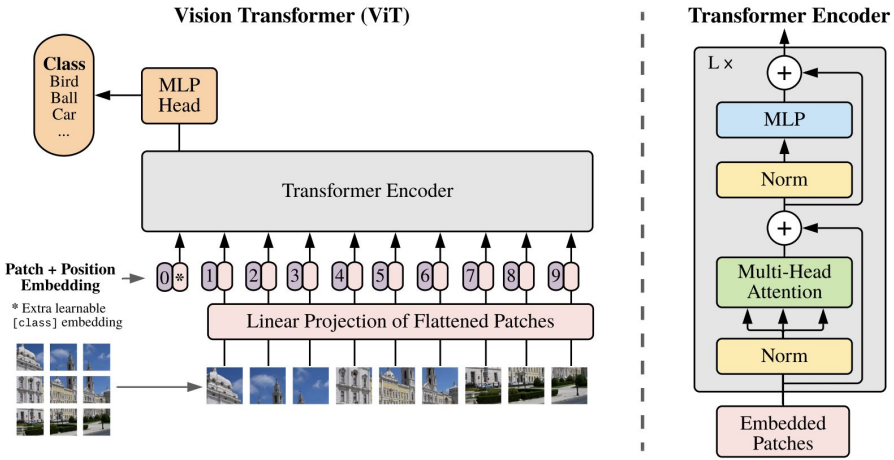
… to attend to **all** positions in the decoder up to and including that position. **We** i
We implement this inside of scaled dot-product **attention** by masking out (setting

☆ 保存 99 引用 被引用数: 98637 関連記事 全 62 バージョン »

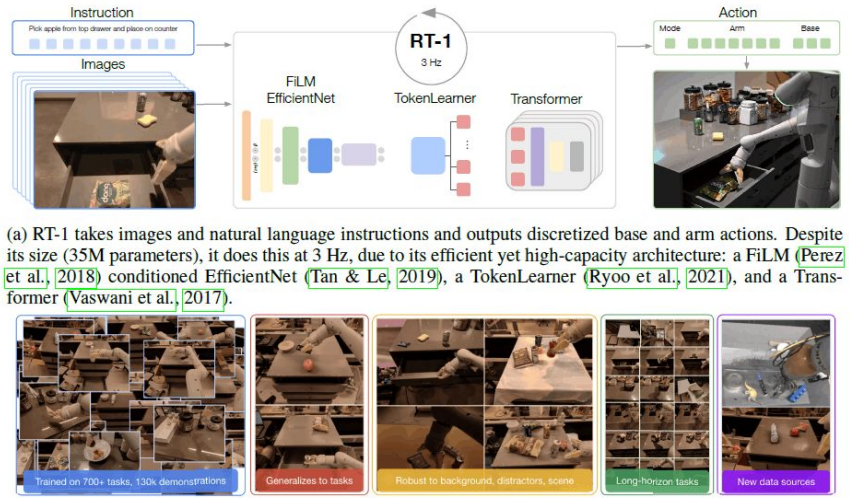
2023/11/30 Google Scholarにアクセス

- アテンション機構を採用することで，トークンの長距離依存関係を効率的に学習できる
- 学習時の並列計算も効率化できたことで大規模化しやすくなった

背景 | Transformerの応用



Vision Transformer



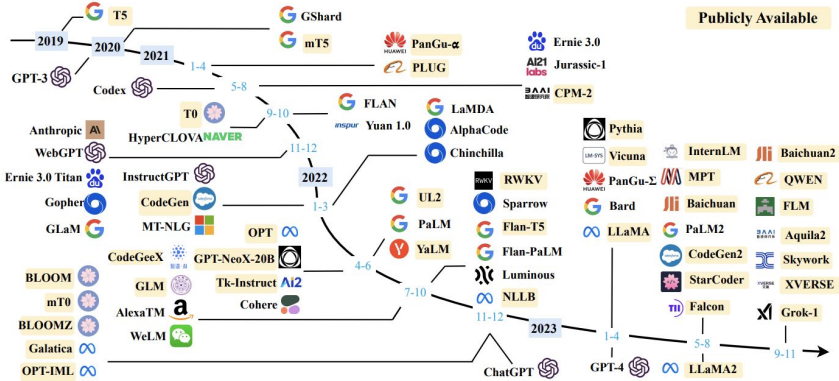
(a) RT-1 takes images and natural language instructions and outputs discretized base and arm actions. Despite its size (35M parameters), it does this at 3 Hz, due to its efficient yet high-capacity architecture: a FiLM (Perez et al., 2018) conditioned EfficientNet (Tan & Le, 2019), a TokenLearner (Ryoo et al., 2021), and a Transformer (Vaswani et al., 2017).

RT-1

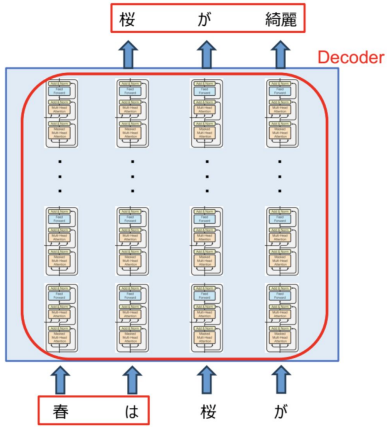
- Vision Transformer: 画像タスク × Transformer
- RT-1: ロボットタスク × Transformer

研究の傾向 | 強化学習 × Transformer

- 画像タスクやロボットタスクなどにTransformerを使用
- 特に言語タスクにおいてTransformerが大成功
  - ChatGPT
- 言語タスクと同様に強化学習も系列データ (軌道) を扱う
  - τ={s,a,r,st+1...}
- 強化学習でもTransformerを活用することができないか？

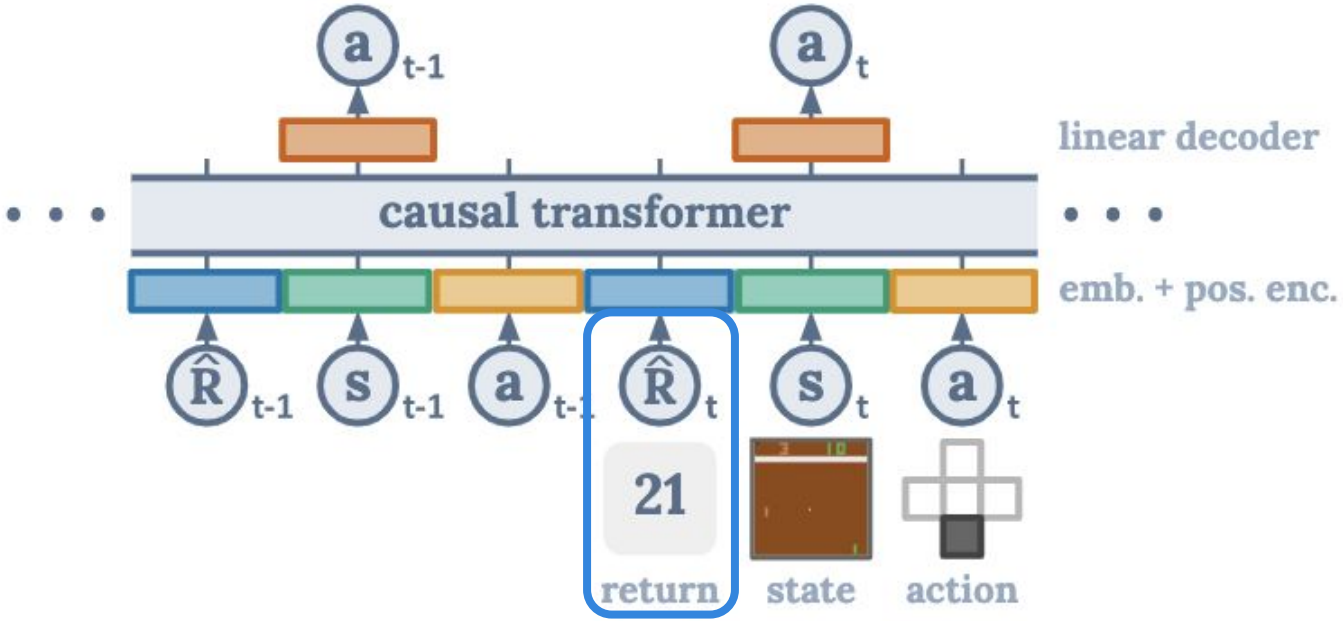


https://arxiv.org/abs/2303.18223



引用：松尾研LLM Summer2023資料

実験設定 | Decision Transformer



- 入力：観測 (s),行動 (a),軌道 (τ) が達成する将来報酬 (Returns-to-go)：t+1以降に獲得する将来報酬和
- 出力：行動 (a) を予測

研究の方向性

Computer Vision and Machine Learning
Kawamoto Lab.

研究テーマ (現在)

Atariゲームに対するTransformerベース強化学習のロバスト性検証

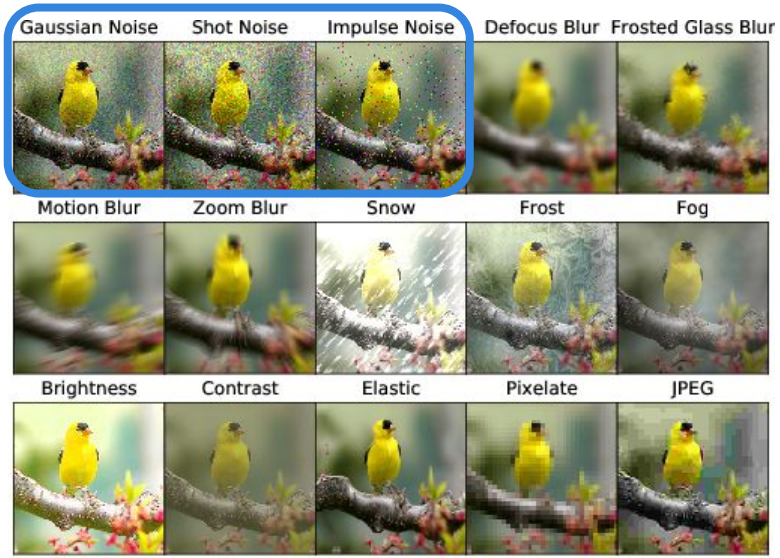
実験設定

Decision Transformer × **Common Corruptions (Noise)**

最終目標

- ❏ Common Corruptionにおける各種ノイズ評価テストにて脆弱性を示す
- ❏ データ拡張訓練により，ロバスト性を向上させる

実験設定 | Common Corruptions (Noise)



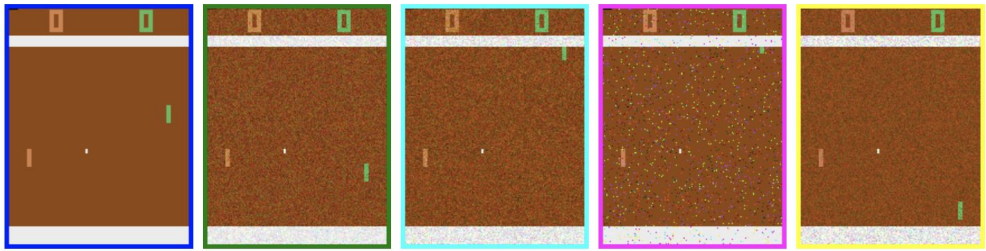


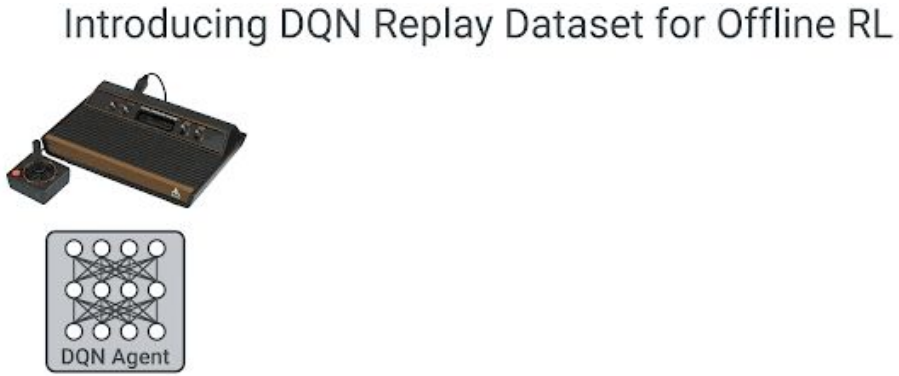
図 12: Pong タスクにおける Clean 訓練に対する各テストの例

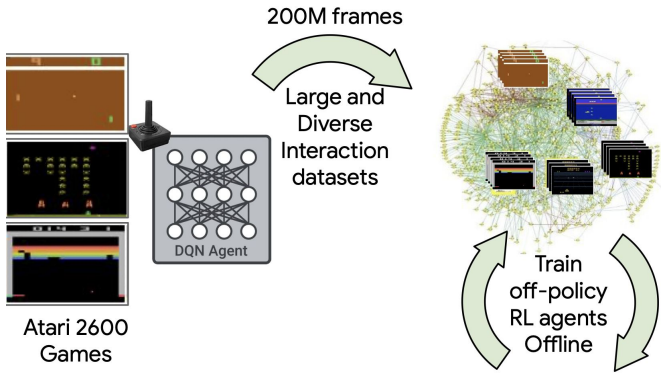
- ❏ Decision Transformerに対するロバスト性検証は確認されていない
- ❏ ゲームタスクへの混入可能性があるノイズ系を対象
  (ぼかし，天候，デジタル系はゲームプレイに混入しない)

実験設定 (補足) | オフラインデータセット

オフライン強化学習は方策から獲得したデータセットを使って学習する

- DQN Replay Dataset
  - 60種類のAtari2600ゲームに対してDQNで訓練した経験
  - 4フレーム (相関を見る) × 5000万タプル (st,at,rt,st+1) = 200M (2億)フレーム
  - 60種類のゲームごとに異なるランダム初期化を行い，5つのデータセットを作成





公式：https://github.com/google-research/batch_rl#dqn-replay-dataset-logged-dqn-data

引用：https://offline-rl.github.io/

研究の方向性

Computer Vision and Machine Learning
Kawamoto Lab.

研究テーマ (現在)

Atariゲームに対するTransformerベース強化学習のロバスト性検証

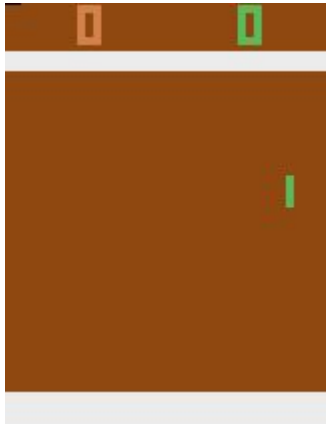
実験設定

Decision Transformer × Common Corruptions (Noise)

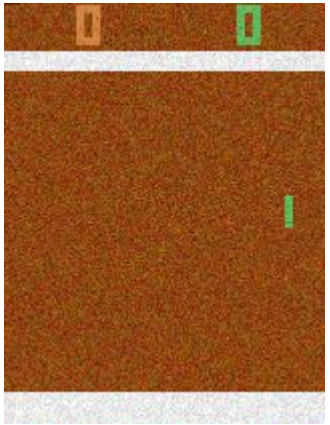
最終目標

- ❑ **Common Corruptionにおける各種ノイズ評価テストにて脆弱性を示す**
- ❑ データ拡張訓練により，ロバスト性を向上させる

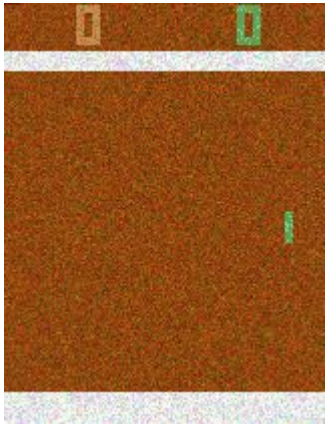
脆弱性検証｜定性評価 (Pong)



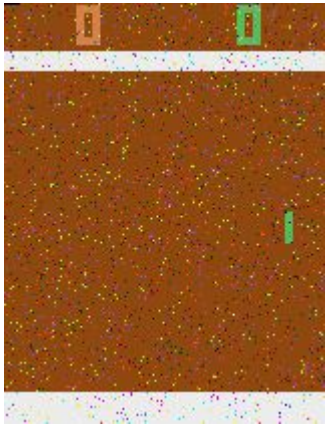
Clean



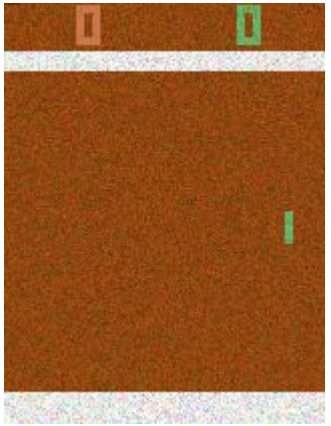
Gaussian



Shot



Impulse



Speckle

- Clean以外のノイズ系テストでは，エージェントが学習しきれていない

脆弱性検証 | 定量評価 (Pong)

表 3: Pong における 3seed 値の平均と標準偏差

	Test				
Train	**Clean**	**Gaussian**	**Shot**	**Impulse**	**Speckle**
Clean	**1.826 ± 17.21**	-20.14 ± 1.035	-19.92 ± 1.233	-20.56 ± 0.7156	-19.99 ± 0.9899
Gaussian	-7.693 ± 17.52	**0.4733 ± 16.07**	-9.033 ± 15.84	-9.813 ± 10.17	1.68 ± 17.22
Shot	1.593 ± 18.02	-1.486 ± 15.68	**-0.5066 ± 16.33**	-12.40 ± 8.369	0.1533 ± 16.96
Impulse	1.0 ± 16.82	-0.3533 ± 15.88	-1.206 ± 15.38	**-1.773 ± 14.84**	0.98 ± 17.04
Speckle	1.746 ± 18.24	-16.86 ± 4.016	-12.16 ± 8.311	-18.32 ± 2.412	**2.34 ± 16.68**

- 通常データで訓練したモデルに対して，テスト時に各種ノイズを付与してプレイすると顕著にゲームスコアが低下する **(ノイズに対して脆弱)**

脆弱性検証 | 定量評価 (Pong)

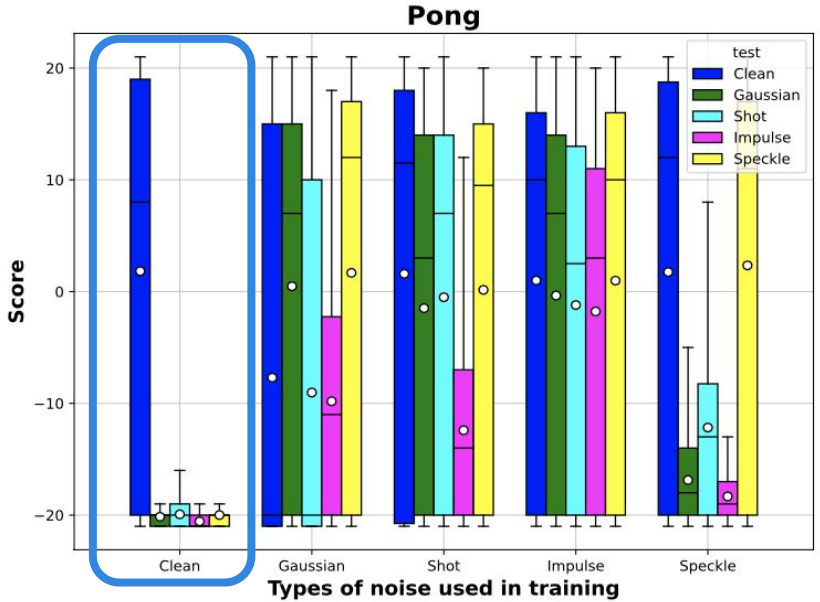


図 6: Pong での各ノイズ訓練に対するスコア

- 通常データで訓練したモデルに対して，テスト時に各種ノイズを付与してプレイするとスコア分布が大幅に低下する

研究の方向性

Computer Vision and Machine Learning
Kawamoto Lab.

研究テーマ (現在)

Atariゲームに対するTransformerベース強化学習のロバスト性検証

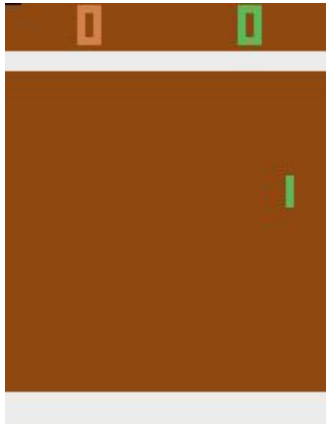
実験設定

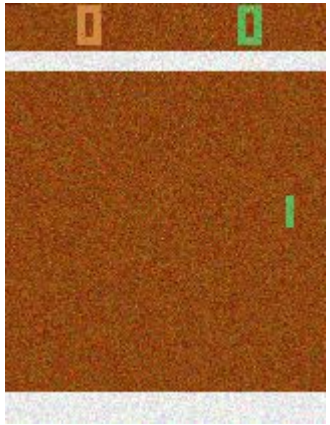
Decision Transformer × Common Corruptions (Noise)

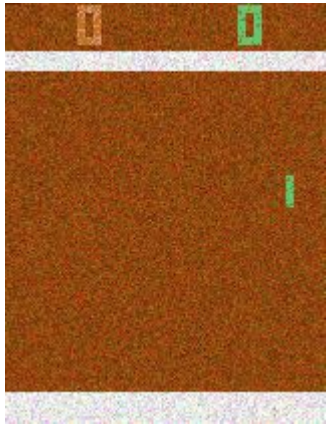
最終目標

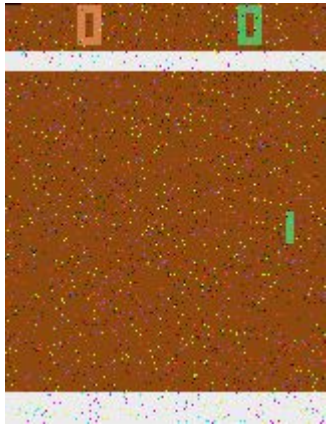
- ❏ Common Corruptionにおける各種ノイズ評価テストにて脆弱性を示す
- ❏ **データ拡張訓練により，ロバスト性を向上させる**

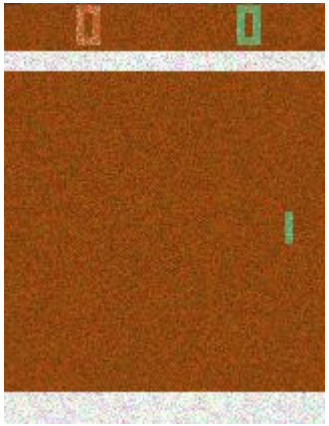
ロバスト性検証 | 定性評価 (Pong)


Clean


Gaussian


Shot


Impulse


Speckle

- 各種訓練データセットに対応するテストの結果
- Clean以外のノイズ系テストでも，エージェントが学習できている

ロバスト性検証 | 定量評価 (Pong)

Computer Vision and Machine Learning Kawamoto Lab.

表 3: Pong における 3seed 値の平均と標準偏差

	Test				
Train	**Clean**	**Gaussian**	**Shot**	**Impulse**	**Speckle**
Clean	**1.826 ± 17.21**	-20.14 ± 1.035	-19.92 ± 1.233	-20.56 ± 0.7156	-19.99 ± 0.9899
Gaussian	-7.693 ± 17.52	**0.4733 ± 16.07**	-9.033 ± 15.84	-9.813 ± 10.17	1.68 ± 17.22
Shot	1.593 ± 18.02	-1.486 ± 15.68	**-0.5066 ± 16.33**	-12.40 ± 8.369	0.1533 ± 16.96
Impulse	1.0 ± 16.82	-0.3533 ± 15.88	-1.206 ± 15.38	**-1.773 ± 14.84**	0.98 ± 17.04
Speckle	1.746 ± 18.24	-16.86 ± 4.016	-12.16 ± 8.311	-18.32 ± 2.412	**2.34 ± 16.68**

- 各種訓練データセットに対応するテストが最も高いスコアを得られた **(ノイズに対してロバスト)**

※他のゲームタスクではノイズデータセット訓練でスコアが上がる傾向

ロバスト性検証 | 定量評価 (Pong)

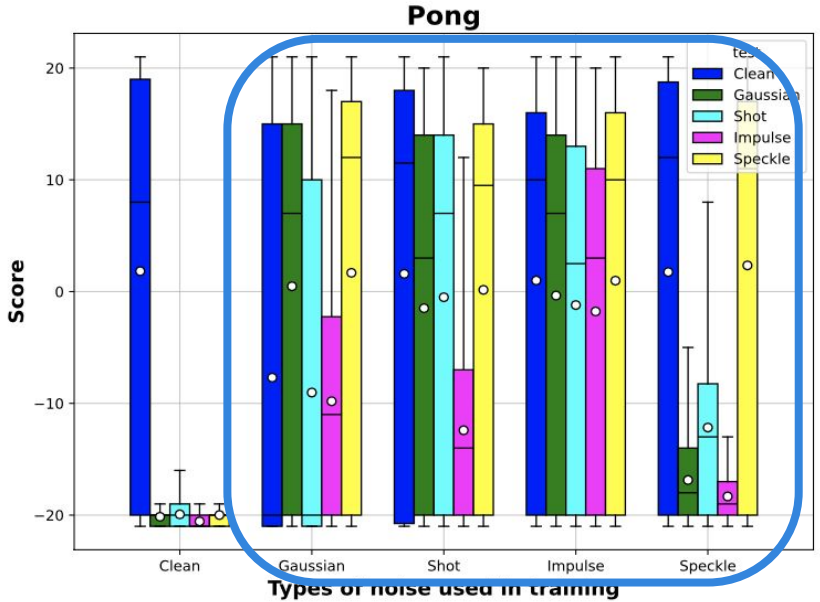


図 6: Pong での各ノイズ訓練に対するスコア

- 通常データ訓練と比較して，各種ノイズデータセット訓練によってスコア分布が向上している

まとめ

- Common Corruptionにおける各種ノイズ評価テストにて脆弱性が存在することを示した
- データ拡張訓練により，ロバスト性が向上することを示した