CSE 525 - Robotics Project Report - 115099704 - Akash G

Introduction

Decision Transformer [1] is a decoder-only transformer architecture that generates actions auto-regressively. It is implemented as a GPT-2 style transformer architecture, whose inputs are K states, K returns-to-go, and (K-1) actions to output the next action (K is the context length or the history of SARs). It can learn from both expert and non-expert data due to reward conditioning.

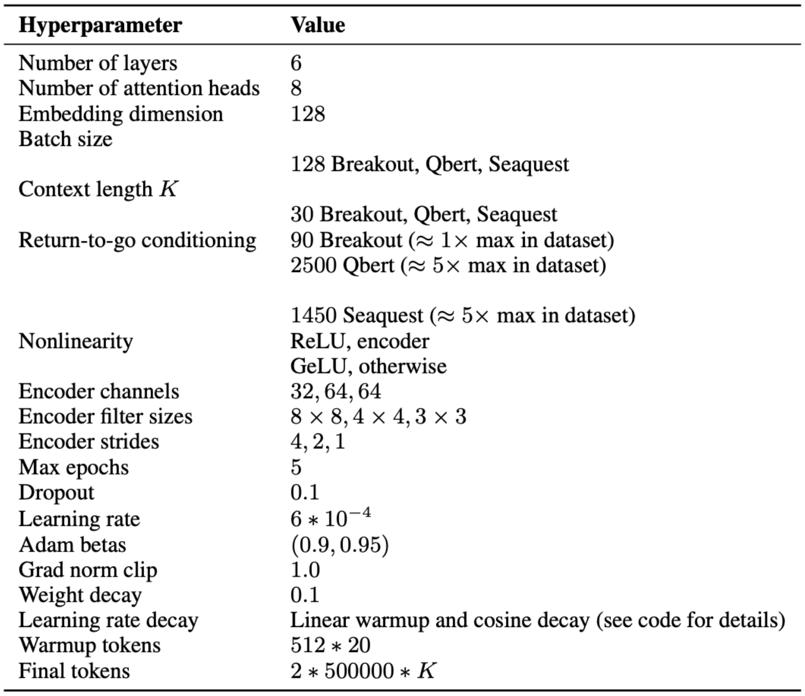
The main motivation for this project is to make use of Vision Transformer [2] to encode the state information for Atari games, in order to improve the performance of the Decision Transformer.

Goals for the Project

1. Train baseline DT, ViT-DT, and MLP BC models on the dataset.
2. Tune CNN, ViT, and MLP hyperparameters for optimal performance.
3. Run for 3 distinct Atari games: Breakout, Seaquest, and Qbert.
4. Average the results over multiple runs with different seeds.

Dataset and Training configurations

DT is primarily trained in an offline learning environment. The agent is trained entirely on an existing dataset [3] without interacting with the environment. The dataset consists of 200M frames or 50M (state, actions, and rewards) experience tuples.



Hyperparameter	**Value**
Number of layers	6
Number of attention heads	8
Embedding dimension	128
Batch size	128 Breakout, Qbert, Seaquest
Context length K	30 Breakout, Qbert, Seaquest
Return-to-go conditioning	90 Breakout (≈ 1× max in dataset) 2500 Qbert (≈ 5× max in dataset) 1450 Seaquest (≈ 5× max in dataset)
Nonlinearity	ReLU, encoder GeLU, otherwise
Encoder channels	$32, 64, 64$
Encoder filter sizes	$8 \times 8, 4 \times 4, 3 \times 3$
Encoder strides	$4, 2, 1$
Max epochs	5
Dropout	0.1
Learning rate	$6 * 10^{-4}$
Adam betas	$(0.9, 0.95)$
Grad norm clip	1.0
Weight decay	0.1
Learning rate decay	Linear warmup and cosine decay (see code for details)
Warmup tokens	$512 * 20$
Final tokens	$2 * 500000 * K$

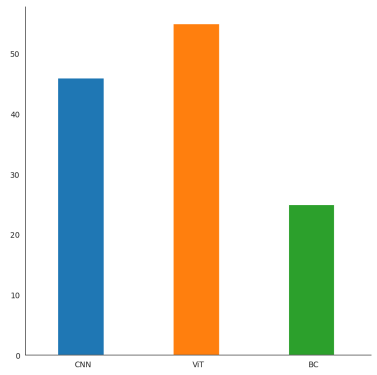
Results

The ViT-based state encoder shows significant improvements in performance over traditional CNN encoders, especially in cases where the attention heads are able to capture varied aspects of the game. For example, in Atari Breakout, each of the attention heads were able to capture the state information of the ball, free-moving space, and the unbroken blocks.

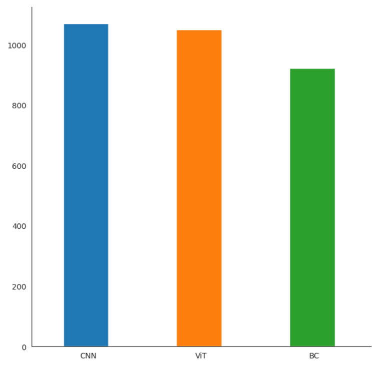
Summary table

Environment	Algorithm	Return To Go	Time taken (hours)	Evaluation reward
Breakout	Baseline DT	90	13.3	46
Breakout	ViT DT	90	45	55
Breakout	Behavior Cloning	90	8.7	25
Seaquest	Baseline DT	2500	6	1071
Seaquest	ViT DT	2500	45	1051
Seaquest	Behavior Cloning	2500	5.5	923
Qbert	Baseline DT	1450	5.8	2450
Qbert	ViT DT	1450	45	5870
Qbert	Behavior Cloning	1450	3.6	2752

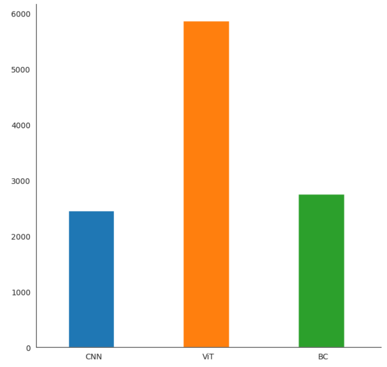
Comparison charts



Breakout
(40% increase)



Seaquest
(2% decrease)



Qbert
(140% increase)

Further developments

1. Multi-game pre-training of DT shows human-level performance: https://arxiv.org/abs/2205.15241 [4]

2. Online fine-tuning approach for DT that learns from replay rollouts (with hindsight experience replay): https://arxiv.org/abs/2202.05607 [5]

3. Prompting DTs to allow better generalization capabilities: https://arxiv.org/abs/2206.13499 [6]

References

[1] Chen, L., Lu, K., Rajeswaran, A., Lee, K., Grover, A., Laskin, M., Abbeel, P., Srinivas, A., & Mordatch, I. (2021). Decision Transformer: Reinforcement Learning via Sequence Modeling. *ArXiv*. /abs/2106.01345

[2] Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., & Houlsby, N. (2020). An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. ArXiv. /abs/2010.11929

[3] Agarwal, Rishabh and Schuurmans, Dale and Norouzi, & Mohammad (2020). An optimistic perspective on offline reinforcement learning. ArXiv. /abs/1907.04543

[4] Lee, K., Nachum, O., Yang, M., Lee, L., Freeman, D., Xu, W., Guadarrama, S., Fischer, I., Jang, E., Michalewski, H., & Mordatch, I. (2022). Multi-Game Decision Transformers. *ArXiv*. /abs/2205.15241

[5] Zheng, Q., Zhang, A., & Grover, A. (2022). Online Decision Transformer. *ArXiv*. /abs/2202.05607

[6] Xu, M., Shen, Y., Zhang, S., Lu, Y., Zhao, D., Tenenbaum, J. B., & Gan, C. (2022). Prompting Decision Transformer for Few-Shot Policy Generalization. *ArXiv*. /abs/2206.13499

[7] J. Shang, X. Li, K. Kahatapitiya, Y. -C. Lee and M. S. Ryoo, "StARformer: Transformer with State-Action-Reward Representations for Robot Learning," in IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, doi: 10.1109/TPAMI.2022.3204708.