Efficient World Models with Context-Aware Tokenization

March 12, 2025

Reference

Vincent Micheli, Eloi Alonso, and François Fleuret. Efficient world models with context-aware tokenization, 2024. URL https://arxiv.org/abs/2406.19320.

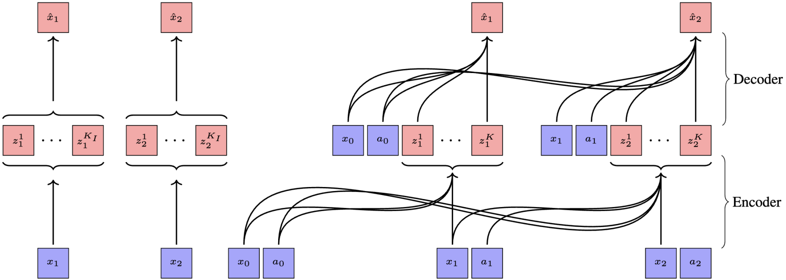
Background

- IRIS agent achieved strong results in Atari 100k benchmark
  - World model: discrete autoencoder + autoregressive transformer
  - Dynamics learning as sequence modelling of image tokens
  - Opened avenues for model-based methods leveraging generative modelling advances
- Scaling challenges:
  - Complex environments require many tokens to encode frames
  - Sophisticated dynamics need longer memory of past states
  - Imagination procedure becomes prohibitively slow
  - Hard to maintain good imagined-to-collected data ratio
- Approach: Δ-IRIS(Micheli et al. [2024])
  - Attends to trajectory of observations and actions
  - Encodes stochastic deltas between time steps
  - Reduces token count and offloads deterministic aspects to autoencoder
  - Interleaves continuous state summaries with discrete transition tokens

Differential Tokens

- Key features of Δ-IRIS:
  - Scales to visually complex environments with longer time horizons
  - Encodes frames by attending to trajectory of observations and actions
  - Describes stochastic deltas between time steps
- Benefits of differential encoding:
  - Drastically reduces number of tokens needed per frame
  - Offloads deterministic aspects to autoencoder
  - Lets transformer focus on stochastic dynamics
- Challenges and solutions:
  - Δ-tokens make autoregressive prediction more difficult
  - Model must integrate over multiple steps to represent current state
  - Solution: Interleave continuous I-tokens (state summaries) with discrete Δ-tokens

IRIS vs Δ-IRIS



Discrete autoencoder comparison: IRIS (left) encodes frames independently, requiring z_t to carry all information for reconstruction. Δ-IRIS (right) conditions on past frames/actions, so z_t only captures stochastic changes. This reduces required tokens ($K \ll K_I$), speeding up autoregressive prediction.

Background: IRIS

- Discrete autoencoder for image tokenization:
  - Encoder maps images to discrete tokens: $E_I : \mathbb{R}^{h \times w \times 3} \rightarrow \{1, \ldots, N_I\}^{K_I}$
  - Decoder reconstructs images from tokens: $D_I : \{1, \ldots, N_I\}^{K_I} \rightarrow \mathbb{R}^{h \times w \times 3}$
  - Trained with reconstruction, perceptual and commitment losses
- Transformer for dynamics modeling:
  - Operates on sequence of image and action tokens
  - Predicts transitions, rewards, and terminations autoregressively
  - Trained with cross-entropy on experience segments
- Key capabilities:
  - Builds reusable vocabulary for frame encoding
  - Attends to history for predictions
  - Models joint distribution of future states

Disentangling deterministic and stochastic dynamics - Part 1

- IRIS limitations:
  - Encodes frames independently - no temporal redundancy assumptions
  - Large token count needed for visually complex frames
  - Quadratic attention cost limits computation
- Solution: Condition autoencoder on history
  - Only encode changes (deltas) between frames
  - Deltas often simpler than full frames
  - Separate deterministic and stochastic components
- Example: Grid-world movement
  - Deterministic: Agent moving based on key press
  - Stochastic: Random enemy appearances
  - Only need to encode stochastic events

Disentangling deterministic and stochastic dynamics- Part 2

- Set definitions:
  - $\mathcal{S}_n(\mathcal{Y}) = \bigcup_{i=1}^{n} \mathcal{Y}^i$ for tuples up to length n
  - $\mathcal{S}(\mathcal{Y}) = \mathcal{S}_\infty(\mathcal{Y})$ for infinite tuples
  - Token vocabulary: $\mathcal{Z} = \{1, \ldots, N\}$
- Encoder $E : \mathcal{S}(\mathcal{X} \times \mathcal{A}) \times \mathcal{X} \to \mathcal{Z}^K$:
  - Input: $(x_0, a_0, \ldots, x_{t-1}, a_{t-1}, x_t)$
  - Output: $z_t = (z_t^1, \ldots, z_t^K)$ Δ-tokens
  - CNN-based with vector quantization
- Decoder $D : \mathcal{S}(\mathcal{X} \times \mathcal{A}) \times \mathcal{Z}^K \to \mathcal{X}$:
  - Input: $(x_0, a_0, \ldots, x_{t-1}, a_{t-1}, z_t)$
  - Output: Reconstructed frame $\hat{x}_t$
  - Losses: $L_1 + L_2 + L_{\text{max}} + L_{\text{commit}}$

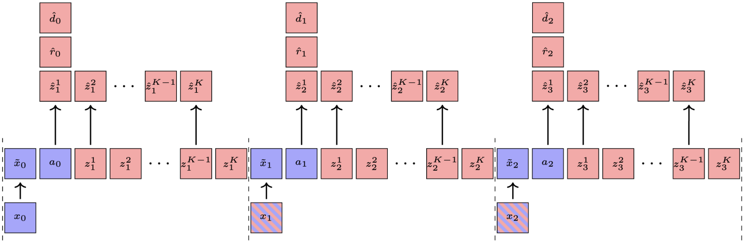
Modelling Stochastic Dynamics - Part 1

- Challenge: Predicting future Δ-tokens is difficult
  - Complex integration over past actions and tokens
  - Example: Random enemy movement in grid world
  - Hard to predict agent-enemy interactions
- Solution: Interleaved continuous I-tokens
  - Similar to MPEG's I-frames
  - Create "soft" Markov blanket
  - Avoid integrating over past Δ-tokens
- I-token Generation:
  - Auxiliary CNN processes frames
  - No discrete autoencoder or reconstruction loss
  - Optimized end-to-end with dynamics model

Modelling Stochastic Dynamics - Part 2

- Dynamics Model Input Sequence:
  - Past I-tokens $\tilde{x}_i$
  - Action tokens a_i
  - Δ-tokens z_i^k
- Model Predictions:
  - Next Δ-token distribution: $p_G(\hat{z}_t^{k+1} | \tilde{x}_{<t}, z_{<t}, a_{<t}, z_t^{\leq k})$
  - Reward distribution: $p_G(\hat{r}_t | \tilde{x}_{\leq t} z_{\leq t}, a_{\leq t})$
  - Termination distribution: $p_G(\hat{d}_t | \tilde{x}_{\leq t}, z_{\leq t}, a_{\leq t})$
- Implementation Details:
  - Transformer encoder with causal self-attention
  - Cross-entropy loss for transitions and terminations
  - Discrete regression with two-hot targets for rewards

Modelling Stochastic Dynamics - Figure



Unrolling dynamics over time. At each step (dashed lines), the GPT-like transformer G predicts Δ-tokens for the next frame, plus reward and termination. It takes action tokens, Δ-tokens, and I-tokens as input, where I-tokens are continuous embeddings that reduce need to attend to past Δ-tokens. Initial frame x_0 embeds to I-token $\tilde{x}_0$. From $\tilde{x}_0$ and a_0, G predicts reward $\hat{r}_0$, termination $\hat{d}_0$, and autoregressively predicts Δ-tokens $\hat{z}_1 = (\hat{z}_1^1, \ldots, \hat{z}_1^K)$. During imagination, next frame (stripped box) is computed by decoder D as $x_1 = D(x_0, a_0, \hat{z}_1)$.

Policy Improvement

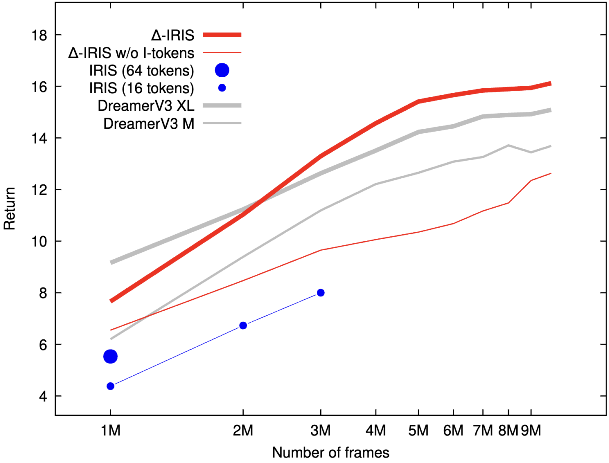
- Learning in imagined POMDP:
  - Policy π trains using world model (E, D, G)
  - Starts from real observation x_0 from experience
  - Rolls out for H steps or until termination
- Imagination procedure:
  - Policy observes reconstructed state: $\hat{x}_t$
  - Samples action: $a_t \sim \pi(a_t|\hat{x}_{\leq t})$
  - Model predicts reward $\hat{r}_t$ and termination $\hat{d}_t$
  - Model generates next tokens: $\hat{z}_{t+1} \sim p_G(\hat{z}_{t+1}|\hat{x}_{\leq t}, \hat{a}_{\leq t}, \hat{z}_{\leq t})$
  - Decoder reconstructs next observation: $\hat{x}_{t+1} = D(\hat{x}_{\leq t}, \hat{a}_{\leq t}, \hat{z}_{\leq t}, \hat{z}_{t+1})$
- Training approach:
  - Actor-critic method from IRIS
  - Value baseline predicts λ-returns
  - REINFORCE with value baseline
  - Entropy maximization for exploration

Experiment

Method	Return @1M	Return @5M	Return @10M	#Parameters	FPS
Δ-IRIS	7.7 (0.5)	**15.4** (0.4)	**16.1** (0.1)	**25M**	20
DreamerV3 XL	**9.2** (0.3)	14.2 (0.2)	15.1 (0.3)	200M	**30**
IRIS (64 tokens)	5.5 (0.7)	-	-	48M	2
Δ-IRIS w/o I-tokens	6.6 (0.2)	10.4 (0.5)	12.6 (0.8)	24M	22
DreamerV3 M	6.2 (0.5)	12.6 (0.7)	13.7 (0.8)	37M	40
IRIS (16 tokens)	4.4 (0.1)	-	-	50M	6

Results on Crafter benchmark: Δ-IRIS achieves best performance at 5-10M frames with 8x fewer parameters than DreamerV3 XL, demonstrating strong scaling in a visually complex environment.

Experiment



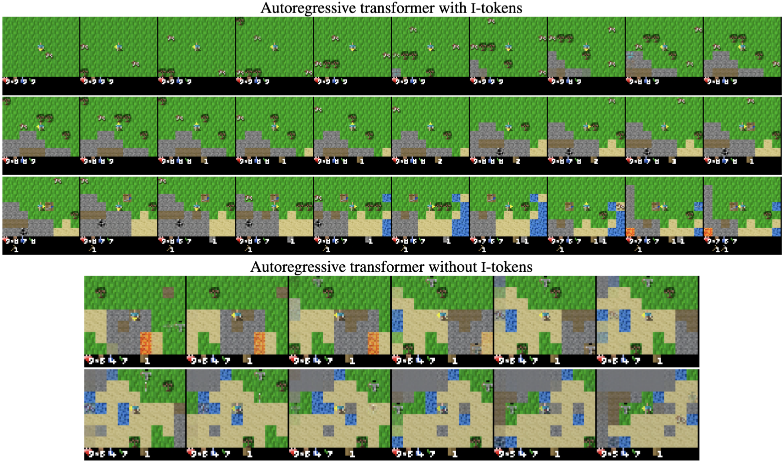
Removing I-tokens from the input sequence of the autoregressive transformer significantly hurts performance.

Reconstruction Errors

Δ-IRIS 4 tokens IRIS 16 tokens

Bottom 1% test frames autoencoded by Δ-IRIS (4 tokens) and IRIS (16 tokens). Each token takes a value in $\{1, 2, \ldots, 1023, 1024\}$, i.e. Δ-IRIS encodes frames with $4 \times \log_2(1024) = 40$ bits while IRIS uses 160 bits. Original frames, reconstructions, and errors are respectively displayed in the top, middle, and bottom rows. Even in the worst instances, Δ-IRIS makes only minor errors, whereas IRIS fails to accurately reconstruct frames. These errors severely hamper the agent's performance, as it purely learns behaviours from frames generated by its autoencoder.

I-tokens



Trajectories imagined with (top) and without (bottom) I-tokens. The top trajectory shows 30+ seconds of coherent gameplay with complex mechanics learned by Δ-IRIS' world model. Without I-tokens, the model fails to predict future Δ-tokens accurately, leading to glitches that hinder policy learning in an unrealistic environment.

Returns on Atari 100k

Game	Random	Human	SimPLe	DreamerV3	STORM	IRIS	Δ-IRIS (ours)
Alien	228	7128	617	959	**984**	420	391
Amidar	6	1720	74	139	**205**	143	64
Assault	222	742	527	706	801	**1524**	1123
Asterix	210	8503	1128	932	1028	854	**2492**
BankHeist	14	753	34	649	641	53	**1148**
BattleZone	2360	37188	4031	12250	**13540**	13074	11825
Boxing	0	12	8	78	**80**	70	70
Breakout	2	31	16	31	16	84	**302**
ChopperCommand	811	7388	979	420	**1888**	1565	1183
CrazyClimber	10781	35829	62584	**97190**	66776	59324	57864
DemonAttack	152	1971	208	303	165	**2034**	533
Freeway	0	30	17	0	**34**	31	31
Frostbite	65	4335	237	909	**1316**	259	279
Gopher	258	2413	597	3730	**8240**	2236	6445
Hero	1027	30826	2657	**11161**	11044	7037	7049
Jamesbond	29	303	101	445	**509**	463	309
Kangaroo	52	3035	51	4098	**4208**	838	2269
Krull	1598	2666	2205	7782	**8413**	6616	5978
KungFuMaster	259	22736	14863	21420	**26182**	21760	21534
MsPacman	307	6952	1480	1327	**2674**	999	1067
Pong	-21	15	13	18	11	15	**20**
PrivateEye	25	69571	35	882	**7781**	100	103
Qbert	164	13455	1289	3405	**4523**	746	1444
RoadRunner	12	7845	5641	15565	**17564**	9615	10414
Seaquest	68	42055	683	618	525	661	**827**
UpNDown	533	11693	3350	**9234**	7985	3546	4072
#Superhuman	0	N/A	1	9	10	10	**11**
Mean	0.00	1.00	0.33	1.10	1.27	1.05	**1.39**
Interquartile Mean	0.00	1.00	0.13	0.50	0.64	0.50	**0.65**