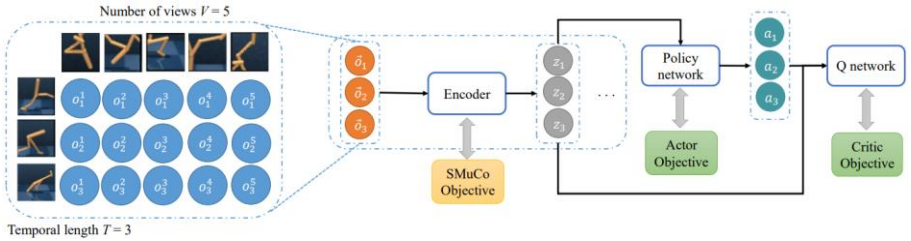
SMuCo: Reinforcement Learning for Visual Control via Sequential Multi-view Total Correlation

Tong Cheng[1], Hang Dong[2], Lu Wang[2], Bo Qiao[2], Qingwei Lin[2], Saravan Rajmohan[3], Thomas Moscibroda[4]

Nanyang Technological University[1], Microsoft AI[2], Microsoft 365[3], Microsoft Azure[4]

Reinforcement Learning for Visual Control

- Importance: leveraging multi-view observation could enhance performance of agent for visual control task
- Challenge: 1) compress task-relevant information; 2) remove task-irrelevant information
- Solution: maximizing sequential multi-view total correlation



Sequential Multi-view Total Correlation

- Extend Multi-view Total Correlation into SMTC
- Derive Lower bound of SMTC

$$SMTC(\vec{O}_{1:T}; Z_{1:T} \mid A_{1:T}) = \sum_{v=1}^{V} I(O^v_{1:T}; Z_{1:T} \mid A_{1:T}) - I(\vec{O}_{1:T}; Z_{1:T} \mid A_{1:T})$$

$$SMTC(\vec{O}_{1:T}; Z_{1:T} \mid A_{1:T}) \geq \sum_{v=1}^{V}\sum_{t=1}^{T} \left[H(O^v_t \mid Z_{t-1}, A_{t-1}) + \mathbb{E}_{p(z_t, o^v_t \mid z_{t-1}, a_{t-1})} \right.$$

$$\left. \ln q^v_\psi(o^v_t \mid z_t, z_{t-1}, a_{t-1}) \right] - \sum_{t=1}^{T}\sum_{s=1}^{T} \mathbb{E}_{p(\vec{o}_s)} \left[D_{KL}(p(z_t \mid o_s, \iota) \parallel r_\phi(z_t \mid \iota)) \right].$$

Training Encoder

- Formulate surrogate function
- Product of Expert (PoE)

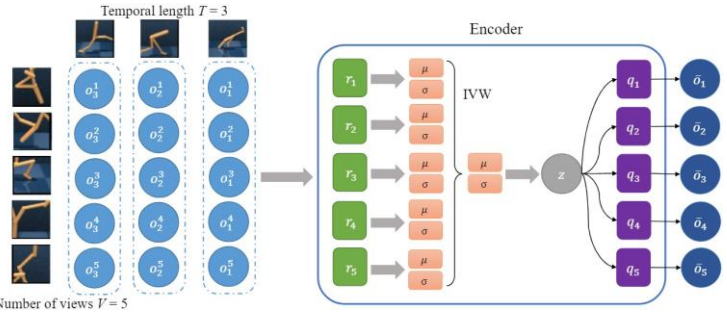
- Inverse Variance Weighted method (IVW)

$$\mathcal{L} = \mathcal{L}_{\text{REC}} + \mathcal{L}_{\text{LL}} + \mathcal{L}_{\text{TC}}$$

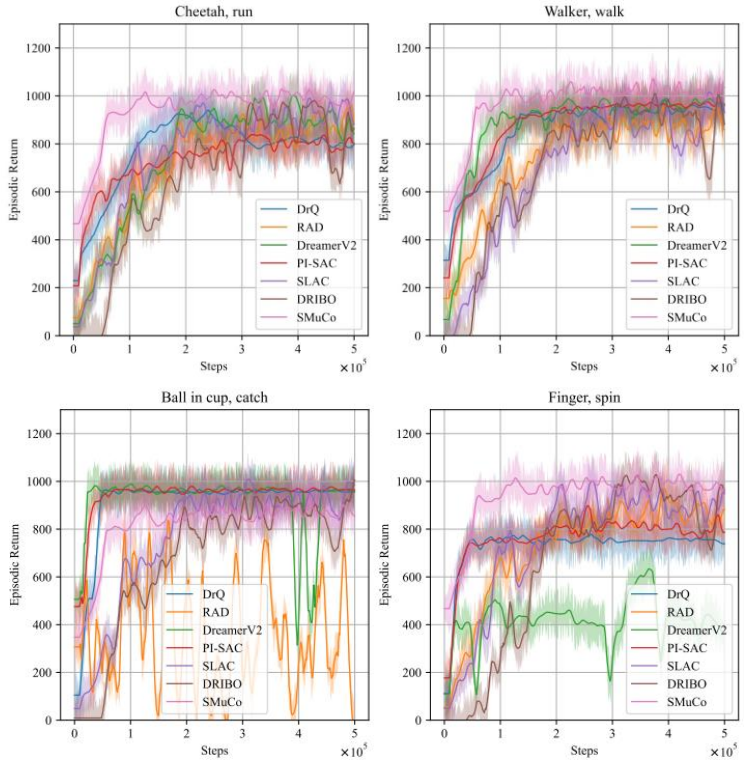
$$\mathcal{L}_{\text{REC}} = -\sum_{v=1}^{V}\sum_{t=1}^{T} H(O^v_t \mid Z_{t-1}, A_{t-1}),$$

$$\mathcal{L}_{\text{LL}} = -\sum_{v=1}^{V}\sum_{t=1}^{T} \mathbb{E}_{p_1} \ln q^v_\psi(o^v_t \mid z_t, z_{t-1}, a_{t-1}),$$

$$\mathcal{L}_{\text{TC}} = \sum_{t=1}^{T}\sum_{s=1}^{T} \mathbb{E}_{p_2}[D_{KL}(p(z_t \mid o_s, \iota) \parallel r_\phi(z_t \mid \iota))],$$



Experiment Results



Scores at 500k Steps	DrQ	RAD	DreamerV2	PI-SAC	SLAC	DRIBO	SMuCo
Cheetah, run	797 ± 116	880 ± 104	841 ± 57	802 ± 119	881 ± 116	864 ± 52	$\mathbf{1019 \pm 107}$
Walker, walk	930 ± 46	858 ± 82	966 ± 117	959 ± 103	930 ± 107	881 ± 90	$\mathbf{1036 \pm 30}$
Ball in cup, catch	958 ± 102	-9 ± 80	955 ± 82	963 ± 75	983 ± 98	$\mathbf{1006 \pm 39}$	853 ± 115
Finger, spin	738 ± 79	880 ± 32	366 ± 105	787 ± 112	947 ± 58	960 ± 53	$\mathbf{969 \pm 52}$
Acrobot, swingup	228 ± 50	163 ± 34	209 ± 106	246 ± 85	$\mathbf{256 \pm 45}$	242 ± 67	247 ± 51
Humanoid, run	470 ± 60	375 ± 117	436 ± 104	482 ± 37	453 ± 117	497 ± 40	$\mathbf{507 \pm 105}$
Hopper, hop	454 ± 89	357 ± 44	347 ± 52	431 ± 67	453 ± 41	$\mathbf{488 \pm 38}$	484 ± 33
Fish, swim	729 ± 90	546 ± 114	697 ± 109	694 ± 102	750 ± 101	$\mathbf{787 \pm 38}$	748 ± 114
Basic Manipulation (uh)	130 ± 30	310 ± 34	108 ± 39	168 ± 45	246 ± 34	84 ± 42	$\mathbf{377 \pm 45}$
Basic Manipulation (ud)	59 ± 34	$\mathbf{366 \pm 32}$	41 ± 38	198 ± 38	242 ± 42	93 ± 43	221 ± 41
Basic Manipulation (hd)	68 ± 45	105 ± 31	78 ± 33	173 ± 32	177 ± 33	70 ± 48	$\mathbf{358 \pm 42}$
Basic Manipulation (uhd)	86 ± 47	368 ± 36	62 ± 40	47 ± 33	136 ± 30	101 ± 38	$\mathbf{446 \pm 34}$