香 港 大 學
THE UNIVERSITY OF HONG KONG

International Conference
On Machine Learning

CtrlFormer: Learning Transferable State Representation for Visual Control via Transformer

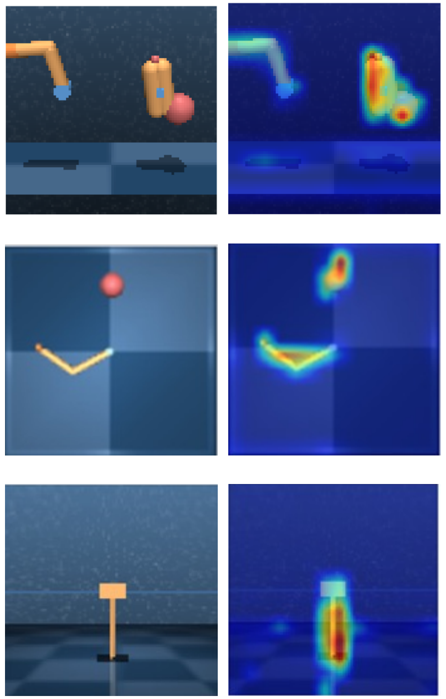
Yao Mu, Shoufa Chen, Mingyu Ding, Jianyu Chen, Runjian Chen, Ping Luo

The University of Hong Kong, Tsinghua University

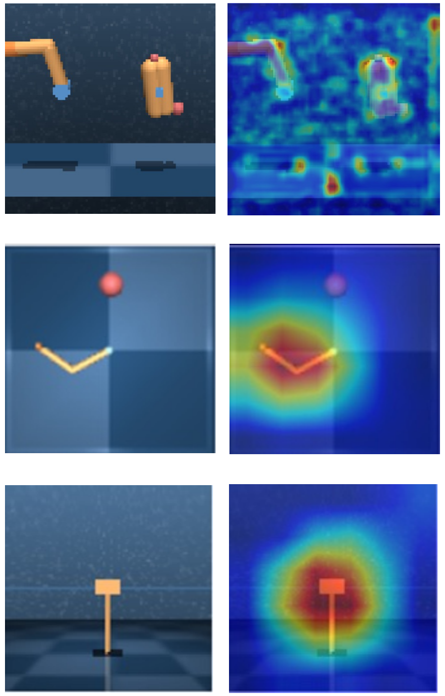
2022.06.27

HKU MMLAB

1 Motivation



Visualization of the attention of task specific visual representation



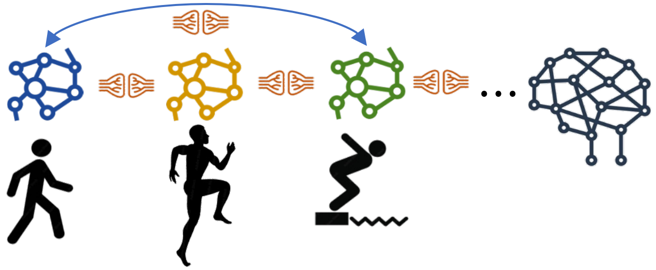
Visualization of the attention of Pretrained ResNet

- Task specific
- High sample efficiency for RL learning
- Difficult to transfer across tasks

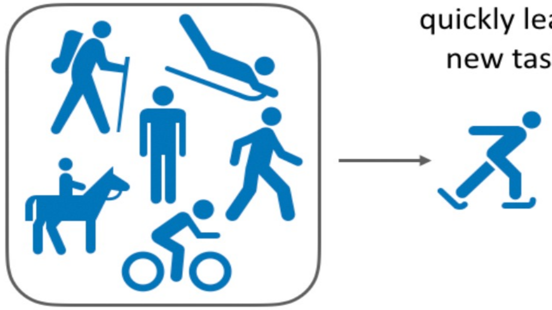
- Task-independent
- Easy to transfer across tasks
- Low sample efficiency for RL learning

Each task has its own focus

Can we learn the visual representation mechanism like a human, which can capture the characteristic of every task and can be easily transferred to a new task?

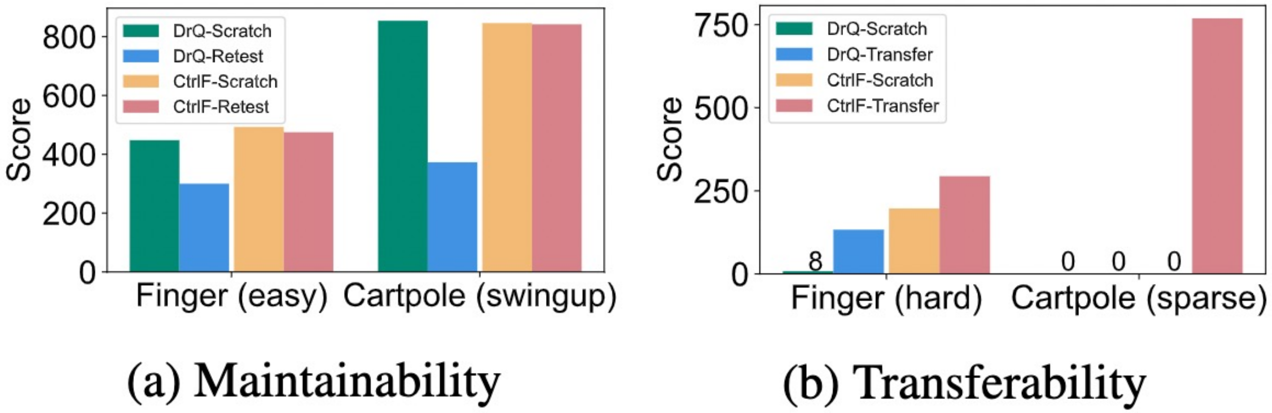


Human behavior learning



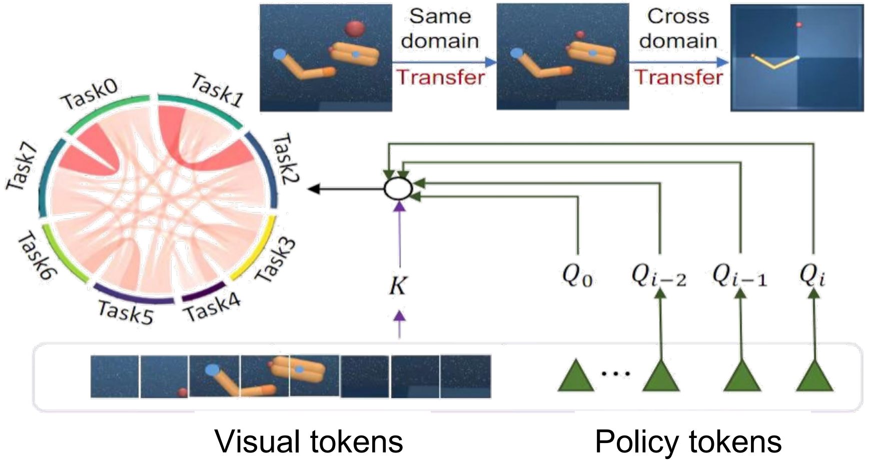
2 Contributions

- CtrlFormer jointly learns self-attention mechanisms between visual tokens and policy tokens among different control tasks, where multitask representation can be learned and transferred without catastrophic forgetting.
- We carefully design a contrastive reinforcement learning paradigm to train CtrlFormer, enabling it to achieve high sample efficiency, which is important in control problems.
- Extensive experiments show that CtrlFormer outperforms previous works in terms of both transferability and sample efficiency without catastrophic forgetting.



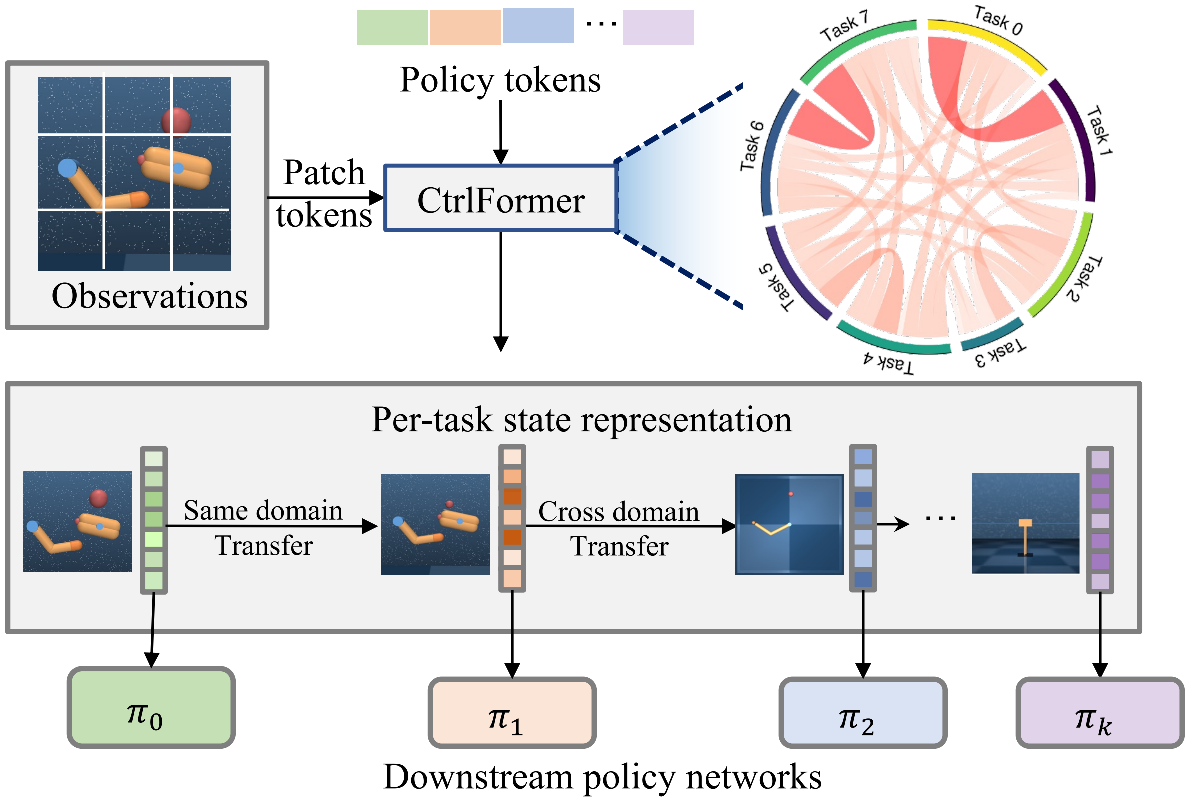
(a) Maintainability (b) Transferability

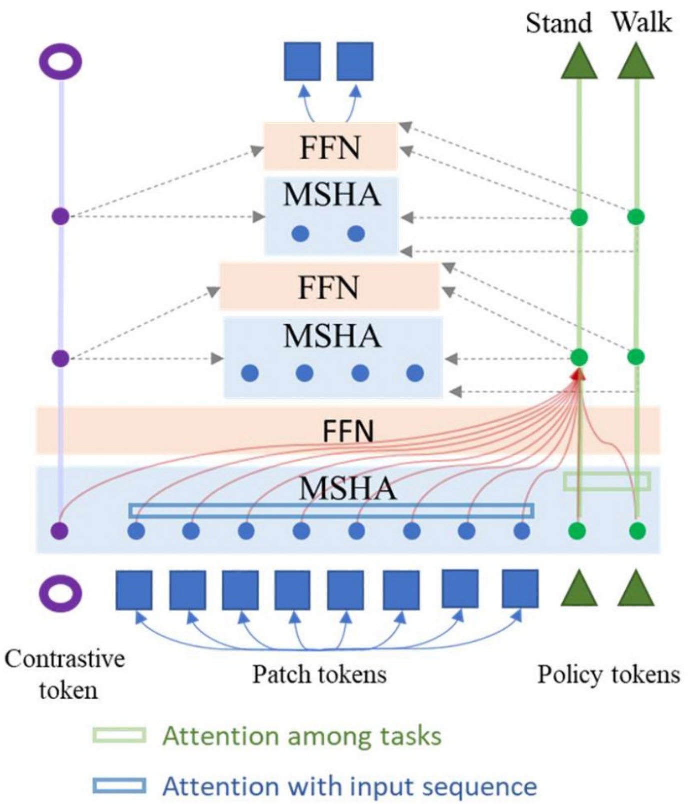
Effect of CtrlFormer



Each task has its own policy token

Overall framework of CtrlFormer

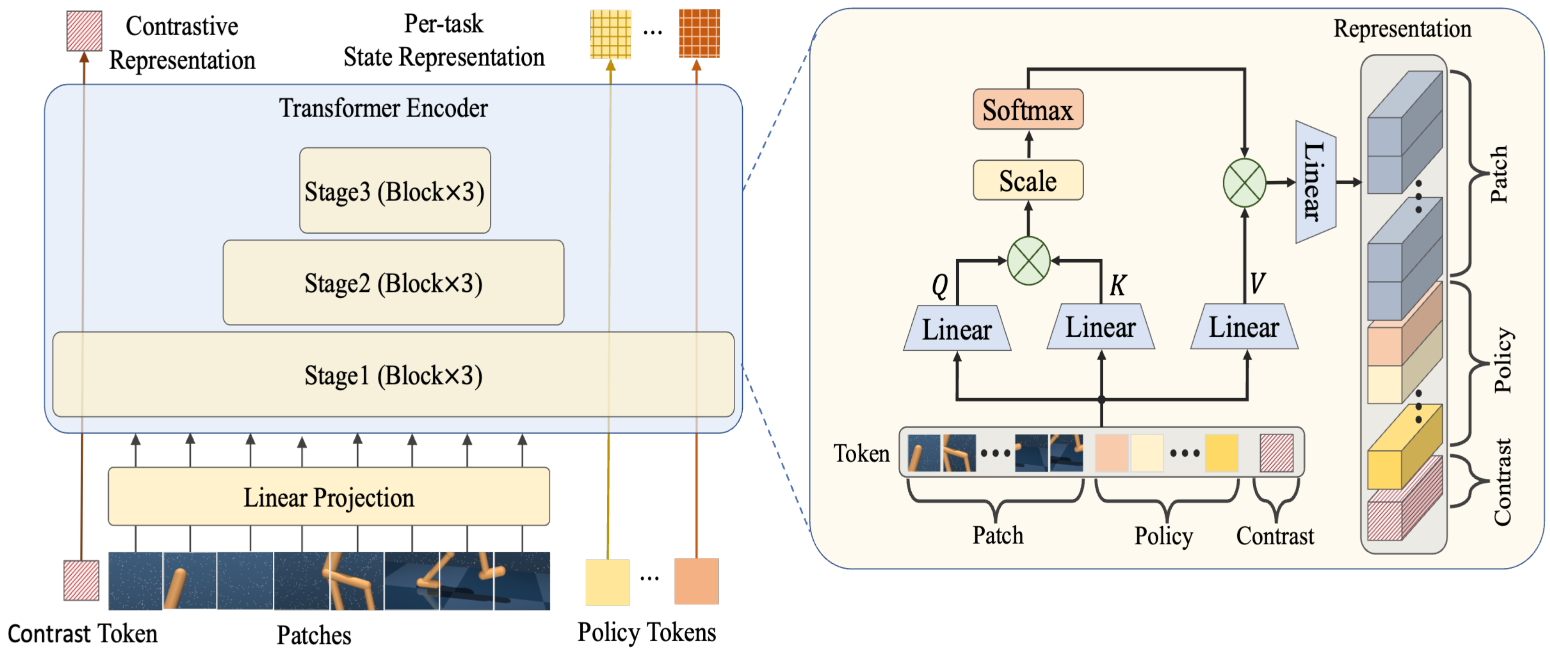




- Policy token is a learnable variable that learns a context for its task during the learning process
- Task-related information can be extracted by computing the attention of policy tokens and other tokens.

Explicitly model the attention mechanism between the new task and the old task and input images thus enabling fast transfer of knowledge learned from the old task to the new one

Detailed structure of CtrlFormer



Thus, the input of the transformer is

$$\mathbf{z}_{\ell_0} = \left[\mathbf{x}_{\text{con}}; \mathbf{x}_\pi^1; \ldots; \mathbf{x}_\pi^K; \mathbf{x}_p^1; \cdots; \mathbf{x}_p^N\right] + \mathbf{E}_{\text{pos}}$$

$$\mathbf{z}'_{\ell_j} = \text{MHSA}\left(\text{LN}\left(\mathbf{z}_{\ell_{j-1}}\right)\right) + \mathbf{z}_{\ell_{j-1}}$$

$$\mathbf{z}_{\ell_j} = \text{MLP}\left(\text{LN}\left(z'_{\ell_j}\right)\right) + \mathbf{z}'_{\ell_j}$$

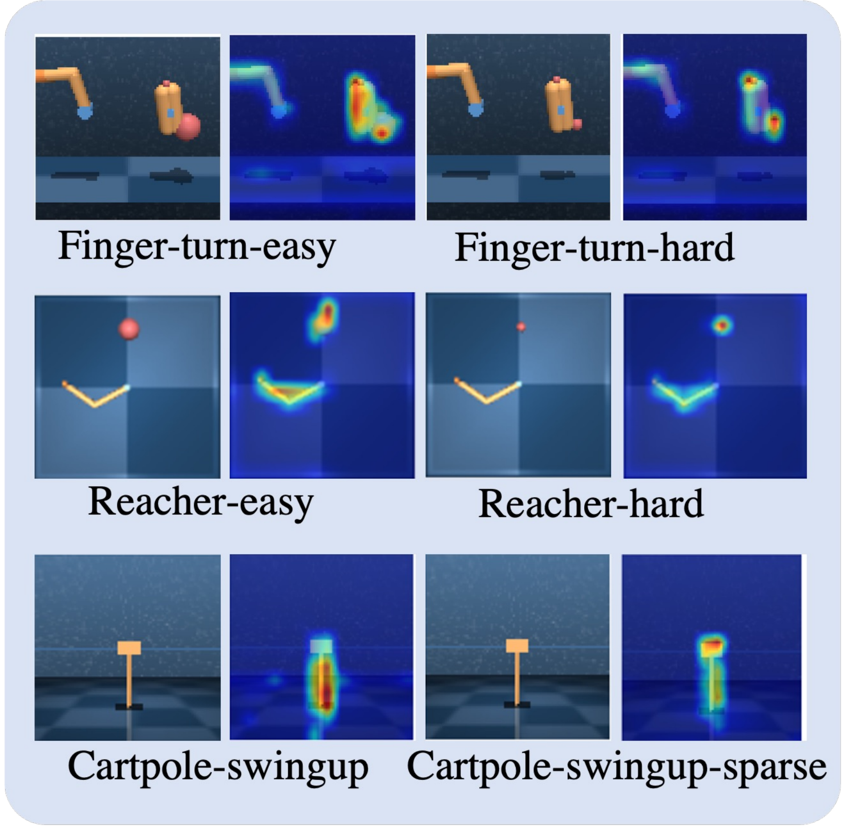
$$\mathbf{q}_\pi = \mathbf{z}_\pi \mathbf{W}^q, \mathbf{k} = \mathbf{z}\mathbf{W}^k$$

$$\mathbf{z} = \left[\mathbf{z}_{con}; \mathbf{z}_\pi^1; \ldots; \mathbf{z}_\pi^{K+1}; \mathbf{z}_p^1; \ldots; \mathbf{z}_p^N\right]$$

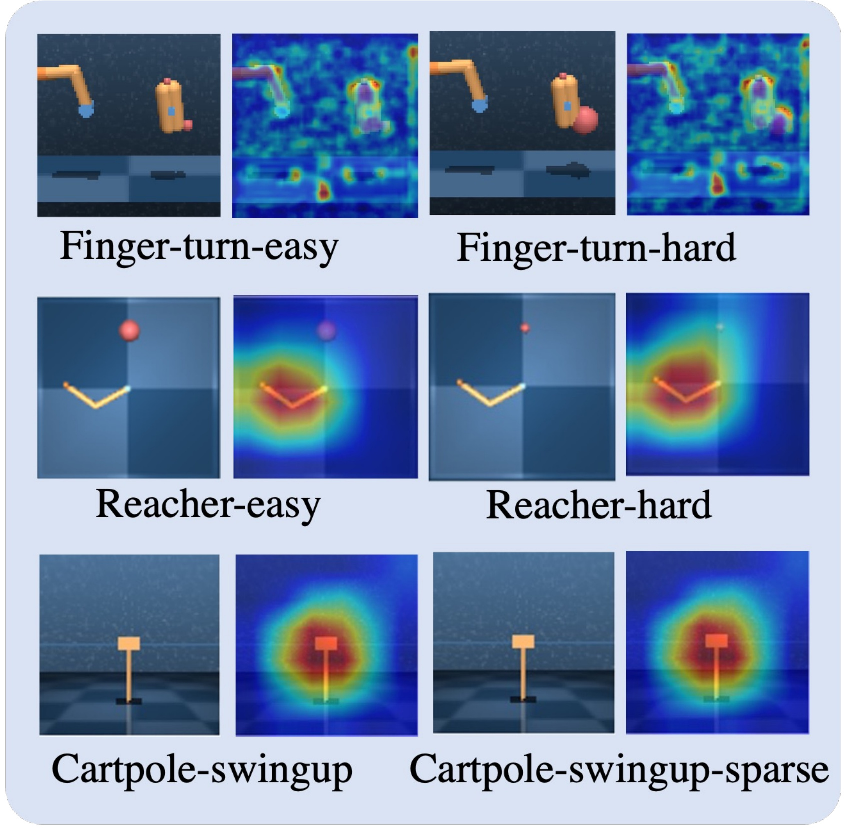
$$\mathbf{z}_\pi = \left[\mathbf{z}_\pi^1; \ldots; \mathbf{z}_\pi^{K+1}\right]$$

- Contrastive Co-training to improve the sample efficiency
- Reduce the number of parameters to be learned by multi-stage Pooling

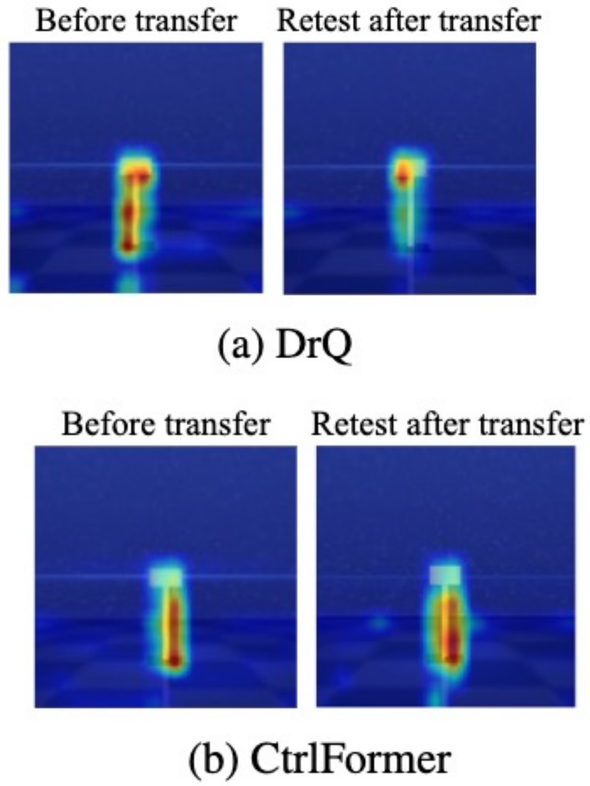
6 Visualization



Visualization of CtrlFormer



Visualization of Pretrained ResNet



Comparison of the attention map change before and after the transferring

7 Results

Method	Learn from scratch 100k	500k	Retest after new task fine-tune
DrQ	$549_{\pm 36}$	$\mathbf{854}_{\pm 22}$	$373_{\pm 24}$
Dreamer	$326_{\pm 27}$	$762_{\pm 27}$	$704_{\pm 33}$
Resnet+SAC	$192_{\pm 19}$	$357_{\pm 85}$	$357_{\pm 85}$
CtrlFormer	$\mathbf{759}_{\pm 48}$	$846_{\pm 25}$	$\mathbf{842}_{\pm 22}$

Method	Learn from scratch 100k	500k	Learn with transfer 100k	500k
DrQ	0	$505_{\pm 335}$	0	$75.5_{\pm 41}$
Dreamer	$8_{\pm 4}$	$376_{\pm 214}$	0	$589_{\pm 122}$
Resnet+SAC	0	0	0	0
CtrlFormer	0	$\mathbf{671}_{\pm 81}$	$\mathbf{769}_{\pm 34}$	$\mathbf{804}_{\pm 26}$

(a) **Left:** Learn old task in Cartpole (`swingup`) **Right:** Transfer to new task Cartpole (`swingup-sparse`)

Method	Learn from scratch 100k	500k	Retest after new task fine-tune
DrQ	$\mathbf{346}_{\pm 33}$	$448_{\pm 65}$	$300_{\pm 42}$
Dreamer	$25_{\pm 18}$	$245_{\pm 159}$	$182_{\pm 34}$
Resnet+SAC	$298_{\pm 17}$	$300_{\pm 29}$	$300_{\pm 29}$
CtrlFormer	$281_{\pm 67}$	$\mathbf{493}_{\pm 35}$	$\mathbf{475}_{\pm 43}$

Method	Learn from scratch 100k	500k	Learn with transfer 100k	500k
DrQ	$8_{\pm 24}$	$274_{\pm 137}$	$133_{\pm 26}$	$455_{\pm 34}$
Dreamer	0.0	$17_{\pm 9}$	0.0	$38_{\pm 18}$
Resnet+SAC	0.0	$17_{\pm 10}$	0.0	$17_{\pm 10}$
CtrlFormer	$\mathbf{197}_{\pm 78}$	$\mathbf{344}_{\pm 47}$	$\mathbf{294}_{\pm 37}$	$\mathbf{569}_{\pm 32}$

(b) **Left:** Learn old task in Finger (`turn-easy`) **Right:** Transfer to new task Finger (`turn-hard`)

Method	Learn from scratch 100k	500k	Retest after new task fine-tune
DrQ	$558_{\pm 38}$	$971_{\pm 27}$	$243_{\pm 52}$
Dreamer	$314_{\pm 155}$	$793_{\pm 164}$	$485_{\pm 67}$
Resnet+SAC	$322_{\pm 285}$	$382_{\pm 299}$	$382_{\pm 299}$
CtrlFormer	$\mathbf{642}_{\pm 42}$	$\mathbf{973}_{\pm 53}$	$\mathbf{906}_{\pm 31}$

Method	Learn from scratch 100k	500k	Learn with transfer 100k	500k
DrQ	$\mathbf{194}_{\pm 84}$	$616_{\pm 274}$	$96_{\pm 43}$	$524_{\pm 68}$
Dreamer	$13_{\pm 32}$	$115_{\pm 98}$	$63_{\pm 07}$	$148_{\pm 12}$
Resnet+SAC	$26_{\pm 4}$	$31.3_{\pm 12}$	$26_{\pm 4}$	$31_{\pm 12}$
CtrlFormer	104 ± 48	$\mathbf{548}_{\pm 131}$	$\mathbf{147}_{\pm 44}$	$\mathbf{657}_{\pm 68}$

(c) **Left:** Learning old task in Reacher (`easy`) **Right:** Transfer to new task Reacher (`hard`)

Method	Learn from scratch 100k	500k	Retest after new task fine-tune
DrQ	$875_{\pm 76}$	$\mathbf{973}_{\pm 65}$	$698_{\pm 57}$
Dreamer	$583_{\pm 21}$	$974_{\pm 31}$	$912_{\pm 19}$
Resnet+SAC	$177_{\pm 32}$	$190_{\pm 24}$	$190_{\pm 24}$
CtrlFormer	$\mathbf{877}_{\pm 42}$	$954_{\pm 38}$	$\mathbf{950}_{\pm 42}$

Method	Learn from scratch 100k	500k	Learn with transfer 100k	500k
DrQ	$504_{\pm 191}$	$\mathbf{947}_{\pm 101}$	$321_{\pm 54}$	$947_{\pm 36}$
Dreamer	$277_{\pm 12}$	$897_{\pm 49}$	$851_{\pm 44}$	$949_{\pm 22}$
Resnet+SAC	$63_{\pm 7}$	$148_{\pm 12}$	$63_{\pm 7}$	$148_{\pm 12}$
CtrlFormer	$\mathbf{593}_{\pm 52}$	$903_{\pm 43}$	$\mathbf{857}_{\pm 47}$	$\mathbf{959}_{\pm 42}$

(d) **Left:** Learning old task in Walker (`stand`) **Right:** Transfer to new task Walker (`walk`)

Method	Scratch (previous) 500 k	Transfer (new task) 100 k	500 k	Retest (previous) 500 k
DrQ	$\mathbf{971}_{\pm 27}$	$283_{\pm 121}$	$332_{\pm 96}$	$124_{\pm 22}$
Resnet+SAC	$382_{\pm 299}$	$298_{\pm 17}$	$300_{\pm 29}$	$382_{\pm 299}$
CtrlFormer	$918_{\pm 33}$	$\mathbf{299}_{\pm 38}$	$\mathbf{547}_{\pm 56}$	$\mathbf{889}_{\pm 34}$

(a) Transfer from Reacher(`easy`) to Finger(`turn-easy`)

Method	Scratch (previous) 500 k	Transfer (new task) 100 k	500 k	Retest (previous) 500 k
DrQ	$\mathbf{448}_{\pm 65}$	$203_{\pm 87}$	$693_{\pm 282}$	$184_{\pm 57}$
Resnet+SAC	$300_{\pm 29}$	$322_{\pm 285}$	$382_{\pm 299}$	$300_{\pm 29}$
CtrlFormer	$424_{\pm 35}$	$\mathbf{416}_{\pm 117}$	$\mathbf{770}_{\pm 71}$	$\mathbf{409}_{\pm 31}$

(b) Transfer from Finger(`turn-easy`) to Reacher(`easy`)

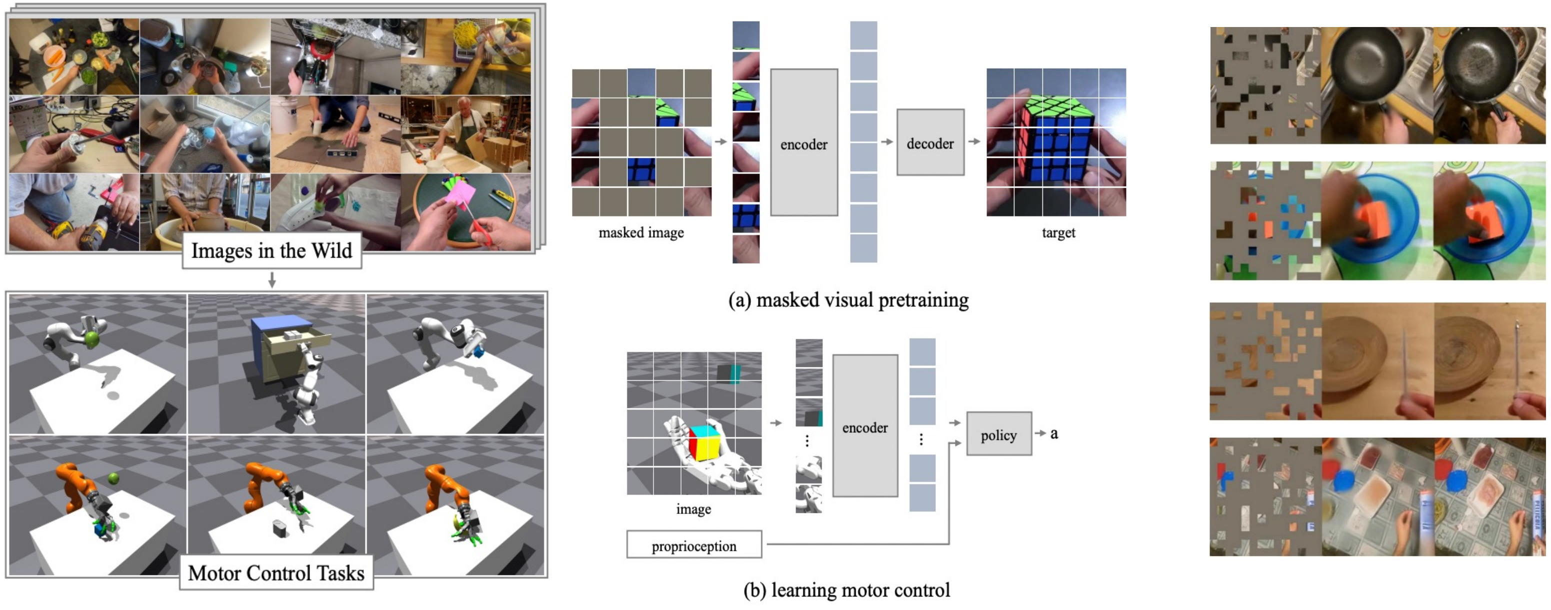
Transfer across multiple tasks

Method	Task 0 $\rightarrow$	Task 1 $\rightarrow$	Task 2 $\rightarrow$	Task 3
Scratch (100k)	$967_{\pm 27}$	$869_{\pm 61}$	$759_{\pm 48}$	0
Train together (100k)	$433_{\pm 23}$	$143_{\pm 34}$	$310_{\pm 41}$	0
CtrlFormer (100k)	$\mathbf{967}_{\pm 27}$	$\mathbf{981}_{\pm 29}$	$\mathbf{988}_{\pm 36}$	$\mathbf{853}_{\pm 69}$
Scratch (500k)	$995_{\pm 18}$	$949_{\pm 44}$	$846_{\pm 25}$	$671_{\pm 81}$
Train together (500k)	$947_{\pm 32}$	$942_{\pm 53}$	$632_{\pm 44}$	$40_{\pm 15}$
CtrlFormer (500k)	$\mathbf{995}_{\pm 18}$	$\mathbf{1000}_{\pm 0}$	$\mathbf{992}_{\pm 26}$	$\mathbf{878}_{\pm 64}$

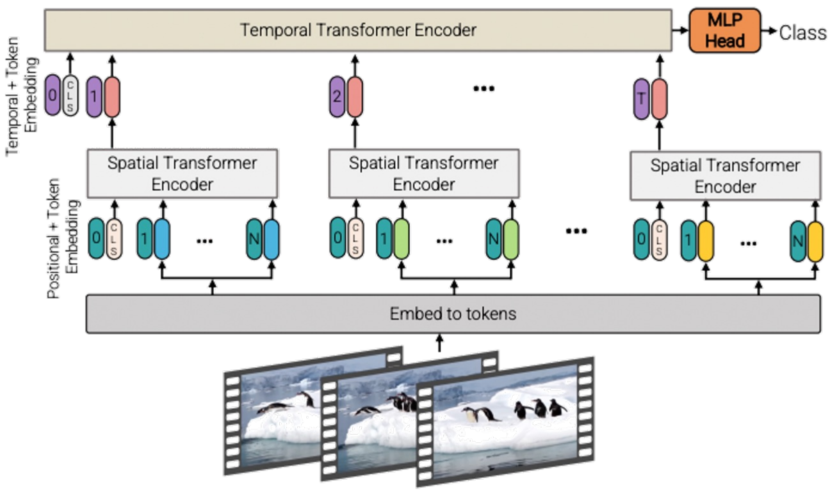
Table 4. **Performance comparison with a series tasks.**

Future works

1. Pretrain the CtrlFormer with the unlabeled data from wild
2. Replace the frame stacking with better temporal modeling



Masked visual pre-training for motor control



Temporal Spatial Transformer

Xiao, Tete, et al. "Masked visual pre-training for motor control." *arXiv preprint arXiv:2203.06173* (2022).

Thanks for Listening! (Q&A)

HKU MMLAB