Elaborating Remark 1: Detailed Comparison of Assumption with Existing Works (review XKke):

Assumption Tag	Related Work	Base Format	Equivalent form
Invariant State Abstraction	Block MDP [Zhang et al., ICML 20']	$T(s'\|s,a;u) = T(\phi(s')\|\phi(s),a) \cdot p_e(u'\|s,u).$	$T(s'\|s,a;u) \propto T(\phi(s')\|\phi(s),a).$
Task independence	Causal Dynamics Learning [Wang et al., ICML 22']	$(\phi(s')\|\phi(s),a) = T(s'_C\|s_C,s_R,a)p(s'_R\|s_R).$	$T(\phi(s^{(1)'})\|\phi(s^{(1)}),a;u^{(1)}) = T(\phi(s^{(2)'})\|\phi(s^{(2)}),a;u^{(2)}).$
Invariant Action Effect	CRL from observational data [Zhu et al., IJCAI 23']	$T_{\text{online}}(s'\|s,a) = \sum_u T(s'\|s,a,u)\hat{p}(u\|s),$ $T_{\text{offline}}(s'\|s,a) = \sum_u T(s'\|s,a,u)\hat{p}(u\|s,a)$	$T_{\text{online}}(s'\|s,a) \propto T_{\text{offline}}(s'\|s,a).$
Invariant Causal Graph	Bilinear MDPs (Ours)	$u^{(1)} \neq u^{(2)}, \mathrm{M}(u^{(1)}) = M(u^{(2)})$	$T(s'\|s,a;u^{(1)}) = T(s'\|s,a;u^{(2)})$

Table I. Comparison to related works in basic assumption. We unify the notation with transition dynamics $T(s'|s,a)$, some state encoder $\phi(s)$ and time-invariant confounded variables $u, u^{(1)} \neq u^{(2)}$.

Additional Baselines: (reviewer ne9G)

Env	MOPO	MOBILE	Ours	Env	MOPO	MOBILE	Ours
Unlock-I-R	21.5 ± 1.9	15.9 ± 1.0	$\mathbf{32.7 \pm 2.8}$	Unlock-O-R	16.6 ± 1.3	12.8 ± 0.8	$\mathbf{27.6 \pm 2.0}$
Unlock-I-M	84.8 ± 5.1	72.4 ± 1.7	$\mathbf{98.0 \pm 4.9}$	Unlock-O-M	39.5 ± 4.7	40.7 ± 1.8	$\mathbf{68.8 \pm 1.5}$
Unlock-I-E	88.8 ± 4.6	78.3 ± 1.2	$\mathbf{94.0 \pm 1.0}$	Unlock-O-E	39.9 ± 4.4	45.6 ± 2.1	$\mathbf{82.1 \pm 6.5}$

Table II. Comparison of the success rate (%) among MOPO, MOBILE (**new!**) and our methods in Unlock environment.

Results on visual observation and inference speed (reviewer 8KHa, XKke):

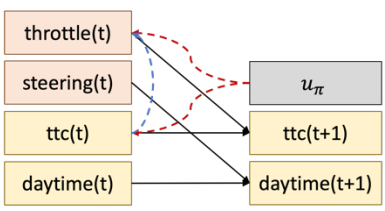
Env	ICIL	IFactor	Ours	Env	ICIL	IFactor	Ours
Unlock-I-R	0.8 ± 0.8	4.3 ± 1.1	$\mathbf{15.7 \pm 3.3}$	Unlock-O-R	1.5 ± 1.8	4.7 ± 1.6	$\mathbf{5.9 \pm .0.9}$
Unlock-I-M	5.3 ± 2.0	30.2 ± 4.1	$\mathbf{62.0 \pm 4.6}$	Unlock-O-M	8.6 ± 4.2	15.4 ± 2.4	$\mathbf{71.6 \pm 9.1}$
Unlock-I-E	8.7 ± 3.4	34.0 ± 4.8	$\mathbf{63.7 \pm 3.9}$	Unlock-O-E	17.1 ± 4.2	16.7 ± 3.1	$\mathbf{73.6 \pm 19.5}$

Table III. Comparison of the success rate (%) among ICIL, IFactor and our methods in Unlock environment with visual observation.

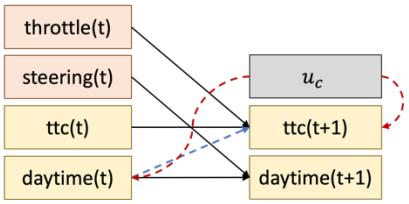
Observation Type	Model-free: ICIL	Model-based: IFactor	Model-based: Ours
Vector	296.0 ± 109.6	32.4 ± 4.0	23.6 ± 4.5
Raw Pixel	106.0 ± 26.2	27.8 ± 1.8	22.5 ± 3.5

Table IV. We report the mean and 95% CI of inference speed (FPS) with model-free and model-based approaches. All the experiments were run on 1-card A6000 GPU, AMD EPYC 7542 32-Core Processor and average over 10 random seeds.

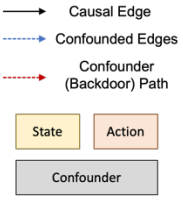
Real-world Example of Causal Confounding Autonomous Driving (reviewer ne9G, 8enR):



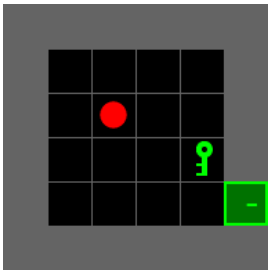
(a) Policy Confounder



(b) Environment Confounder



(c) Visual Observation of *Unlock*



Experiment Details (reviewer ne9G, 8KHa, XKke):

Models	Parameters	Value	Models	Parameters	Value
MOBILE*	n_model	7	IFactor (Visual)*	prior output size	32
	n_elites	5		hidden_dims visual	128
	hidden_dims	128		Layers of CNN	3
Ours (Visual)*	hidden_dims visual	128	ICIL (Visual)	Layers of CNN	3
	Layers of CNN	3		EBM hiddens	128

Table V. Important hyperparameters of MOBILE, and BECAUSE, IFactor, ICIL with visual inputs. (*) means the approach is based on the same set of MPC controller as BECAUSE. with the vector state in training pairs. For visual RL setting, during inference, the vector state will be absent.