Supplementary Material for *Isolating and Leveraging Noncontrollable Visual Dynamics in World Models*

Minting Pan* **Xiangming Zhu*** **Yunbo Wang†** **Xiaokang Yang**
MoE Key Lab of Artificial Intelligence, AI Institute, Shanghai Jiao Tong University
{panmt53, xmzhu76, yunbow, xkyang}@sjtu.edu.cn

This supplementary material consists of five parts, including

- Implementation details on each benchmark (Section 1).
- Descriptions of the compared methods (Section 2).
- More qualitative results in the DMC suite and CARLA simulator (Section 3).
- Ablation studies on the BAIR dataset (Section 4).
- Network details of Iso-Dream for different environments (Section 5).

1 Benchmarks

We quantitatively and qualitatively evaluate Iso-Dream on the following two environments for visual control and two real-world datasets for action-conditioned video prediction.

- **DeepMind control suite** [12]: A set of stable, well-tested continuous control tasks that are easy to use and modify. For vision-based control, we use a modified version of the DeepMind control suite in DMControl Generalization Benchmark [9] to evaluate Iso-Dream. In this environment, agents are trained to complete different tasks with random natural video as backgrounds, namely `video_easy` and `video_hard` benchmarks. We use 4 tasks to test our Iso-Dream, *i.e.*, Finger Spin, Cheetah Run, Walker Walk, Hopper Stand.
- **CARLA** [3]: An open-source simulator with more complex and realistic visual observations for autonomous driving research. In our experiments, we evaluate Iso-Dream in a first-person highway driving task in "Town04". The agent's goal is to drive as far as possible in 1000 time steps without colliding with the 30 other moving vehicles or barriers.
- **BAIR robot pushing** [4]: An action-conditioned video prediction dataset composed of hours of self-supervised learning with the robotic arm Sawyer. In each video, a random moving robotic arm pushes a variety of objects on similar tables with a static background. Each video also has recorded actions taken by the robotic arm which correspond to the commanded gripper pose.
- **RoboNet** [1]: A large-scale dataset contains action-conditioned videos of seven robotic arms interacting with a variety of objects from four different research laboratories, *i.e.*, Berkeley, Google, Penn, and Stanford.

2 Compared Methods

For visual MBRL, we compare our method with the following baselines and existing approaches:

*Equal contribution.
†Corresponding author: Yunbo Wang.
Code available at `https://github.com/panmt/Iso-Dream`

36th Conference on Neural Information Processing Systems (NeurIPS 2022).

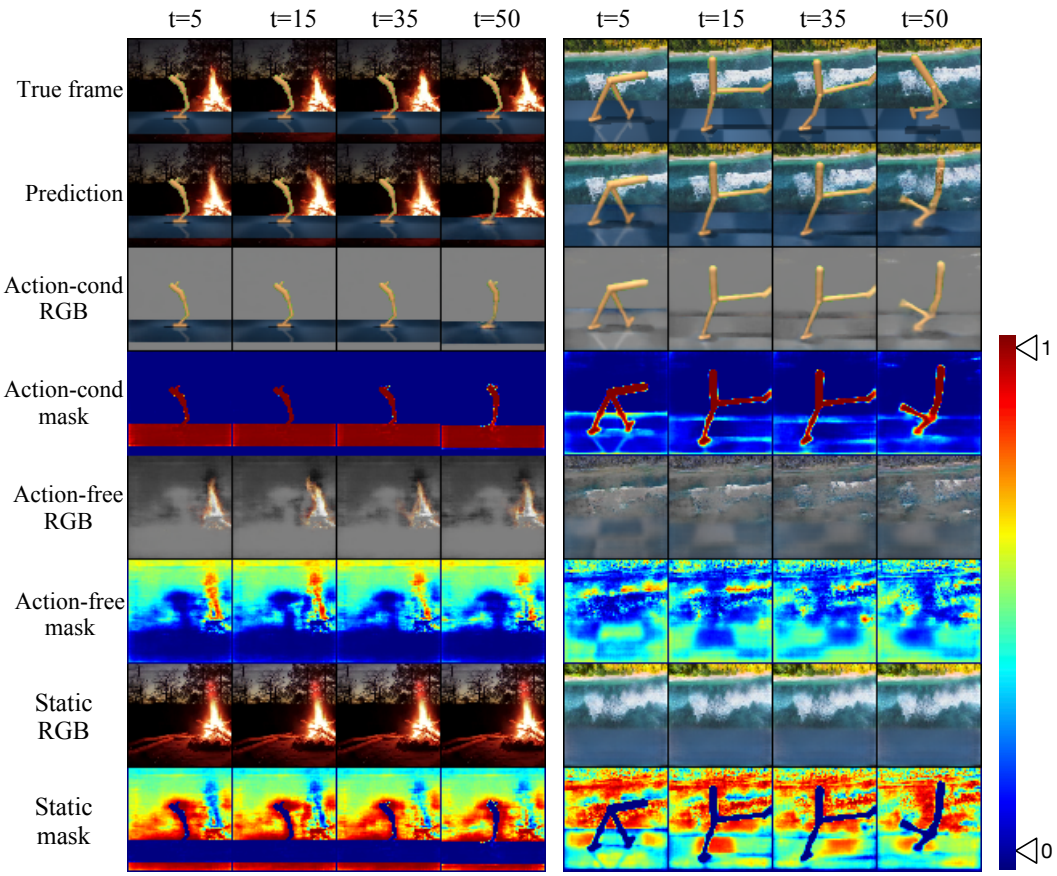

Figure 1: Video prediction results with different noisy backgrounds on the DMC. For each sequence, we use the first 5 images as context frames.

- **DreamerV2** [7]: A model-based RL method that learns directly from latent variables in world models. The latent representation enables agents to imagine thousands of trajectories in parallel.
- **CURL [10]**: A model-free RL method that extracts high-level features from raw pixels using contrastive learning, maximizing agreement between augmented versions of the same observation.
- **SVEA [8]**: A framework for data augmentation in deep Q-learning algorithms that improves stability and generalization on off-policy RL.
- **SAC [6]**: A model-free actor-critic method that optimizes a stochastic policy in an off-policy way.
- **DBC [13]**: It learns a bisimulation metric representation without reconstruction loss, which are invariant to different task-irrelevant details in the observation.

For video prediction, we compare the proposed world model with the following approaches:

- **SVG** [2]: This model introduces random variables into latent space, which ensures that the future trajectory is inherently random.
- **SA-ConvLSTM** [11]: Based on the self-attention mechanism, this model uses the self-attention memory to capture long-term spatial dependency.
- **PhyDNet** [5]: This model uses a two-branch architecture to disentangle PDE dynamics from unknown complementary information.

3 Additional Visualization in DMC and CARLA

DeepMind Control suite. In Figure 1, more showcases on the DeepMind Control are presented with different noisy backgrounds. We show the visualization of the masks and decoupled components from three branches of Iso-Dream.

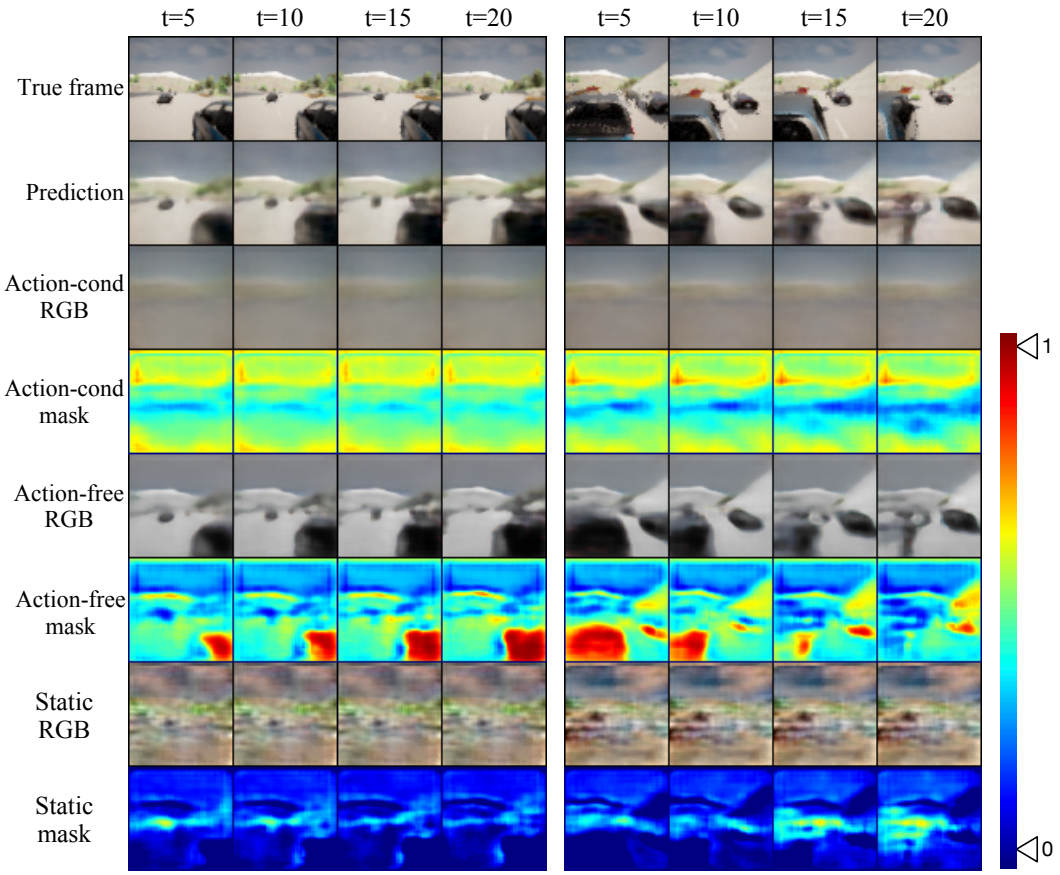

Figure 2: Video prediction results with 10 vehicles (**left**) and 20 vehicles (**right**) on the CARLA environment. For each sequence, we use the first 5 images as context frames.

Table 1: Ablation study for each component of Iso-Dream for video prediction on BAIR with bouncing balls. Lines 1-2 show the results of removing the action-free branch and Inverse cell, respectively. We use the first 2 frames as input to predict the next 18 frames and the next 28 frames.

Model	Predict 18 frames		Predict 28 frames	
	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑
Iso-Dream w/o action-free branch	20.47	0.795	18.51	0.690
Iso-Dream w/o Inverse cell	21.42	0.829	19.34	0.759
Iso-Dream	**21.43**	**0.832**	**19.51**	**0.768**

CARLA autonomous driving simulator. In Figure 2, we visualize the video prediction results on the CARLA environment with different numbers of vehicles. We train Iso-Dream with 30 vehicles and test with 10 vehicles and 20 vehicles respectively.

4 Ablation study on the BAIR Robot Pushing Dataset

In Table 1, the first row shows the results of removing the action-free branch in the world model of Iso-Dream. The performance has decreased from 21.43 to 20.47 and from 19.51 to 18.51 in PSNR for predicting the next 18 frames and next 28 frames respectively, indicating that modular network structures are effective for predictive learning by decoupling the controllable and noncontrollable representations. Comparing the second row and third row in the Table 1, we observe that modeling

Table 2: An overview of layers and hyper-parameters used for three environments.

Name	DMC	CARLA	BARI / RoboNet
Enc_θ	conv3-32	conv3-32	conv3-64
`Action-conditioned branch`			
$Enc_{\phi 1}$	conv3-64	conv3-64	conv3-64
$\texttt{GRU}_s$	hidden size = 200	hidden size = 200	-
ST-LSTM	-	-	hidden size = 64
$Dec_{\varphi 1}$	conv3-4	conv3-4	conv3-4
α	1	1	0.0001
β_1	1	1	-
`Action-free branch`			
$Enc_{\phi 2}$	conv3-64	conv3-64	conv3-64
$\texttt{GRU}_z$	hidden size = 200	hidden size = 200	-
ST-LSTM	-	-	hidden size = 64
$Dec_{\varphi 2}$	conv3-4	conv3-4	conv3-4
β_2	-	1	-
`Static branch`			
$Enc_{\phi 3}$	conv3-64	conv3-64	-
$Dec_{\varphi 3}$	conv3-3	conv3-3	-

inverse dynamics can improve the performance by learning more deterministic state transitions given particular actions in the action-conditioned branch.

5 Network Architectures for Different Environments

The networks and hyper-parameters used for different environments are shown in Table 2.

References

[1] Sudeep Dasari, Frederik Ebert, Stephen Tian, Suraj Nair, Bernadette Bucher, Karl Schmeckpeper, Siddharth Singh, Sergey Levine, and Chelsea Finn. Robonet: Large-scale multi-robot learning. *arXiv preprint arXiv:1910.11215*, 2019.

[2] Emily Denton and Rob Fergus. Stochastic video generation with a learned prior. In *ICML*, pages 1174–1183. PMLR, 2018.

[3] Alexey Dosovitskiy, Germán Ros, Felipe Codevilla, Antonio M. López, and Vladlen Koltun. CARLA: an open urban driving simulator. In *CoRL*, volume 78, pages 1–16. PMLR, 2017.

[4] Frederik Ebert, Chelsea Finn, Alex X Lee, and Sergey Levine. Self-supervised visual planning with temporal skip connections. In *CoRL*, pages 344–356, 2017.

[5] Vincent Le Guen and Nicolas Thome. Disentangling physical dynamics from unknown factors for unsupervised video prediction. In *CVPR*, pages 11474–11484, 2020.

[6] Tuomas Haarnoja, Aurick Zhou, Kristian Hartikainen, George Tucker, Sehoon Ha, Jie Tan, Vikash Kumar, Henry Zhu, Abhishek Gupta, Pieter Abbeel, et al. Soft actor-critic algorithms and applications. *arXiv preprint arXiv:1812.05905*, 2018.

[7] Danijar Hafner, Timothy Lillicrap, Mohammad Norouzi, and Jimmy Ba. Mastering atari with discrete world models. *arXiv preprint arXiv:2010.02193*, 2020.

[8] Nicklas Hansen, Hao Su, and Xiaolong Wang. Stabilizing deep q-learning with convnets and vision transformers under data augmentation. In *NeurIPS*, 2021.

[9] Nicklas Hansen and Xiaolong Wang. Generalization in reinforcement learning by soft data augmentation. In *ICRA*, 2021.

[10] Michael Laskin, Aravind Srinivas, and Pieter Abbeel. CURL: contrastive unsupervised representations for reinforcement learning. In *ICML*, volume 119 of *Proceedings of Machine Learning Research*, pages 5639–5650. PMLR, 2020.

[11] Zhihui Lin, Maomao Li, Zhuobin Zheng, Yangyang Cheng, and Chun Yuan. Self-attention convlstm for spatiotemporal prediction. In *AAAI*, volume 34, pages 11531–11538, 2020.

[12] Yuval Tassa, Yotam Doron, Alistair Muldal, Tom Erez, Yazhe Li, Diego de Las Casas, David Budden, Abbas Abdolmaleki, Josh Merel, Andrew Lefrancq, et al. Deepmind control suite. *arXiv preprint arXiv:1801.00690*, 2018.

[13] Amy Zhang, Rowan McAllister, Roberto Calandra, Yarin Gal, and Sergey Levine. Learning invariant representations for reinforcement learning without reconstruction. In *ICLR*, 2021.