A Code Release

We plan to make the code for Segmentation Dreamer publicly available upon acceptance.

B Visualization of Tasks

B.1 DeepMind Control suite (DMC)

Fig. 6 visualizes the six tasks in DMC (Tassa et al., 2018) used in our experiments. Each row presents the observation from the standard environment, the corresponding observation with added distractions, the ground-truth segmentation mask, and the RGB target with the ground-truth mask applied. Cartpole Swingup Sparse and Cartpole Swingup share the same embodiment and dynamics. Cartpole Swingup Sparse only provides a reward when the pole is upright, whereas Cartpole Swingup continuously provides dense rewards weighted by the proximity of the pole to the upright position. Reacher Easy entails two objects marked with different colors in the segmentation mask, as shown in Fig. 6e 3rd column. Before passing the mask to SD, the mask is converted to a binary format where both objects are marked as *true* as task-relevant.

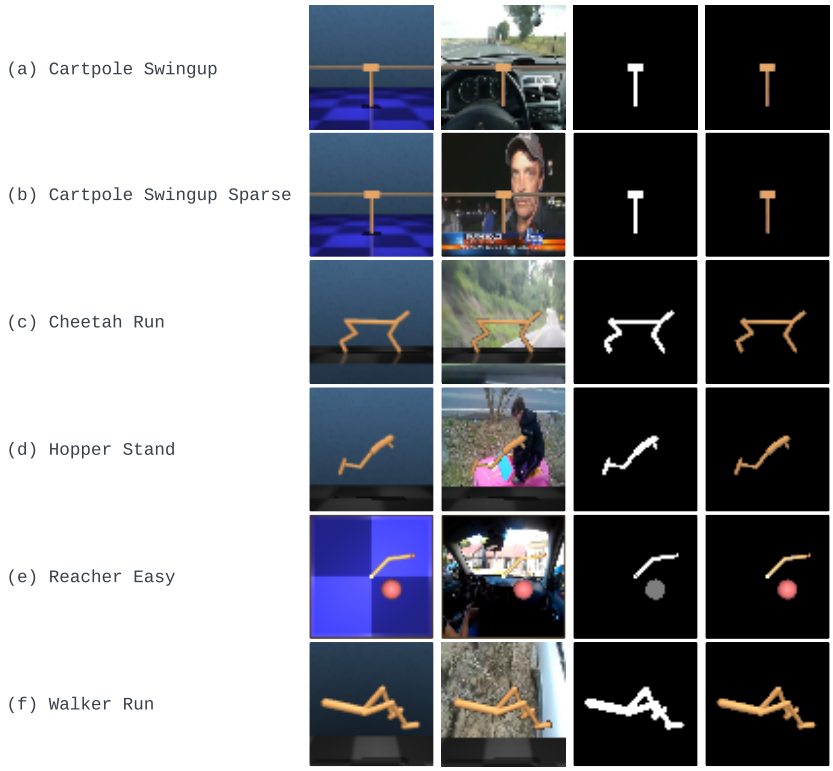


Figure 6: DMC tasks. Left to right: (1) standard environment observations, (2) distracting environment observations, (3) ground-truth segmentation masks, and (4) RGB observations with ground-truth masks applied. We use (4) as auxiliary reconstruction targets in SD^{GT}.

B.2 META-WORLD

Fig. 7 shows the six tasks from Meta-World-V2 used in our experiments. Meta-World is a realistic robotic manipulation benchmark with challenges such as multi-object interactions, small objects, and occlusions.

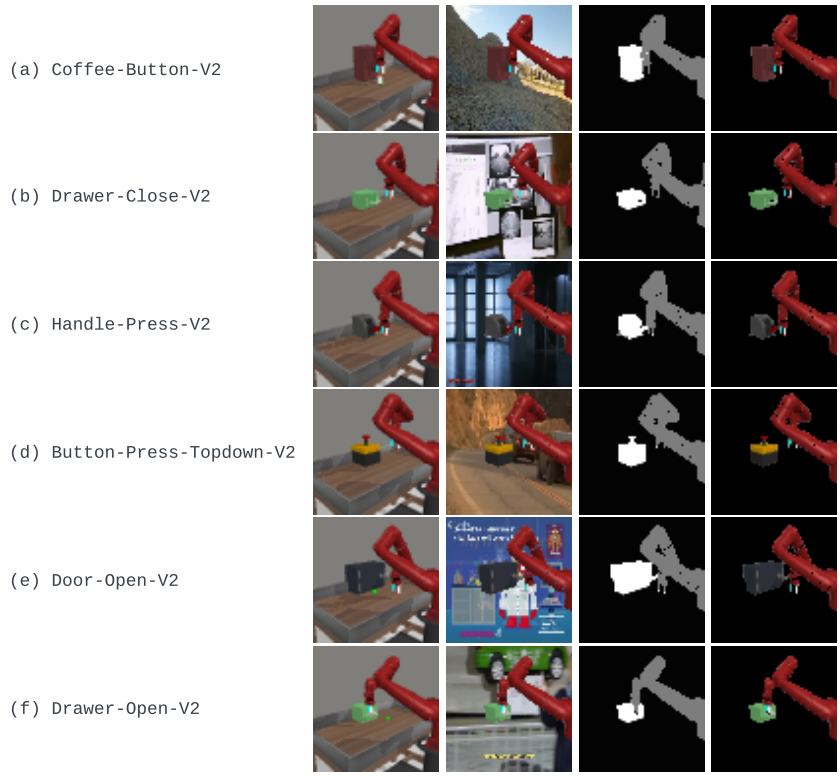


Figure 7: Meta-World tasks. Left to right: (1) standard environment observations, (2) distracting environment observations, (3) ground-truth segmentation masks, and (4) RGB observations with ground-truth masks applied. We use (4) as auxiliary reconstruction targets in SD^{GT}. Masks with multiple classes for different objects are converted to binary masks (all non-background regions are *true* and task-relevant) before use with SD.

C THE IMPACT OF PRIOR KNOWLEDGE

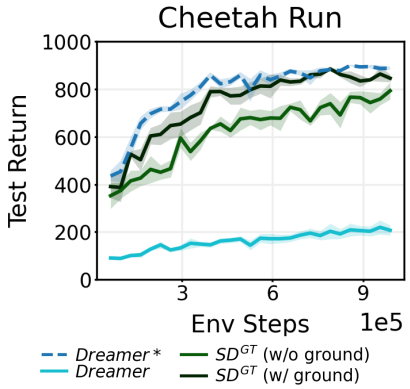
We investigate the impact of accurate prior knowledge of task-relevant objects. Specifically, we conduct additional experiments on Cheetah Run—the task showing the largest disparity between DREAMER* and SD^{GT} in Fig. 3a. In our primary experiment, we designated only the cheetah's body as the task-relevant object. However, since the cheetah's dynamics are influenced by ground contact, the ground plate should have also been considered task-relevant.

Fig. 8 (a–c) illustrates the observation with distractions, the auxiliary target without the ground plate, and with the ground plate included, respectively. Fig. 8d compares SD^{GT} trained with different selections of task-relevant objects included in the masked RGB reconstruction targets. We show that including the ground plate leads to faster learning and performance closer to that of the oracle. This highlights the significant influence of prior knowledge on downstream tasks, suggesting that comprehensively including task-relevant objects yields greater benefits.

(a) Observation Input

(b) Target w/ Ground-Truth **w/o Ground Plate** Applied

(c) Target w/ Ground-Truth **w/ Ground Plate** Applied



(d) Evaluation Return of SD^{GT} with Different Prior Knowledge

Figure 8: **The impact of prior knowledge on Cheetah Run.** (d) The mean over 4 seeds with the standard error of the mean (SEM) is shaded.

D THE IMPACT OF TEST-TIME SEGMENTATION QUALITY ON PERFORMANCE

We investigate how test-time segmentation quality affects $\text{SD}^{\text{approx.}}$ as well as the *As Input* variation that applies mask predictions to RGB inputs in addition to reconstruction targets. For this analysis, we use PerSAM fine-tuned with a single data point for segmentation prediction. To measure segmentation quality, we compute episodic segmentation quality by averaging over frame-level IoU. In Fig. 9 we plot episode segmentation quality versus test-time reward on the evaluation episodes during the last 10% of training time.

Fig. 9 illustrates that $\text{SD}^{\text{approx.}}$ exhibits greater robustness to test-time segmentation quality compared to the *As Input* variation, with the discrepancy increasing as the IoU decreases. This disparity primarily arises because *As Input* relies on observations restricted by segmentation predictions, and thus its performance deteriorates quickly as the segmentation quality decreases. In contrast, $\text{SD}^{\text{approx.}}$ takes the original observation as input and all feature extraction is handled by the observation encoder, informed by our masked RGB reconstruction objective. Consequently, $\text{SD}^{\text{approx.}}$ maintains resilience to test-time segmentation quality.

An intriguing observation is that a poorly trained agent can lead to poor test-time segmentation quality. For instance, Cartpole Swingup (Sparse) exhibits different segmentation quality distributions between $\text{SD}^{\text{approx.}}$ and *As Input*. This discrepancy occurs because the sub-optimal agent often positions the pole at the cart track edge, causing occlusion and hindering accurate segmentation prediction by PerSAM.

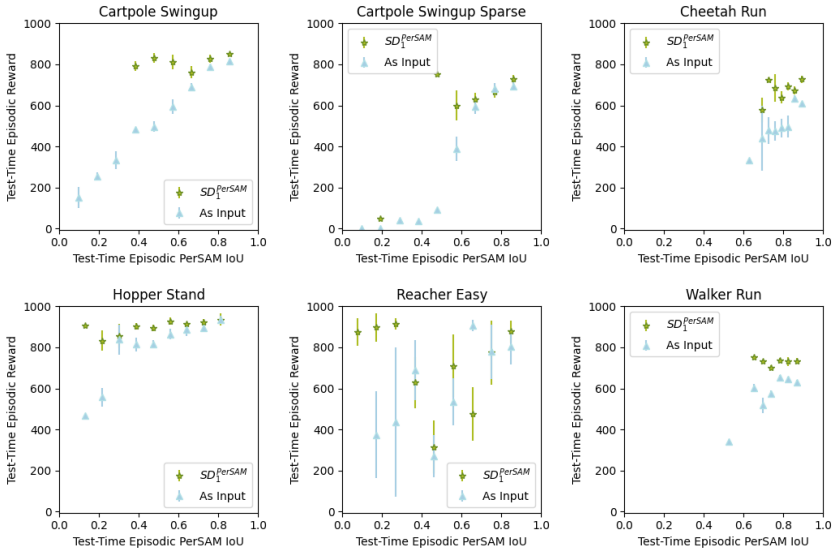


Figure 9: Test-time episodic reward vs PerSAM episodic IoU for $\text{SD}_1^{\text{PerSAM}}$ and *As Input* ($\text{SD}_1^{\text{PerSAM}}$ with masked RGB observations as input). $\text{SD}_1^{\text{PerSAM}}$ is more robust to test-time segmentation prediction errors.

E ABLATION WITHOUT STOP GRADIENT

Should the SD$^{\text{approx.}}$ world model be shielded from gradients of the binary mask decoder head?

To estimate potential regions on RGB targets where task-relevant regions are incorrectly masked out, we train a binary mask prediction head on the world model to help detect false negatives in masks provided by the foundation model. We see better performance when gradients from this binary mask decoder objective are not propagated to the rest of the world model. Thus, the default SD$^{\text{approx.}}$ architecture is trained with the gradients of the binary mask branch stopped at its $[h_t; z_t]$ inputs, and the latent representations in the world model are trained only by the task-relevant RGB branch in addition to the standard DREAMER reward/continue prediction and KL-divergence between the dynamics prior and observation encoder posterior. Tab. 2 shows that the performance drops significantly when training without stopping these gradients.

We also examine masks predicted by the binary mask decoder head in Fig. 10. Predictions are coarser grained than their RGB counterparts, lacking details important for predicting intricate forward dynamics. Overall, reconstructing RGB observations with task-relevance masks applied demonstrates itself as a superior inductive bias to learn useful features for downstream tasks compared to binary masks or raw unfiltered RGB observations.

Table 2: Final performance of SD and SD without stop gradient.

Task	SD$_1^{\text{PerSAM}}$	No SG
Cartpole Swingup	**730** $\pm$ **75**	439 $\pm$ 81
Cartpole Swingup Sparse	**521** $\pm$ **92**	112 $\pm$ 40
Cheetah Run	**619** $\pm$ **35**	376 $\pm$ 50
Hopper Stand	**846** $\pm$ **27**	587 $\pm$ 127
Reacher Easy	**597** $\pm$ **97**	273 $\pm$ 74
Walker Run	**730** $\pm$ **13**	407 $\pm$ 62

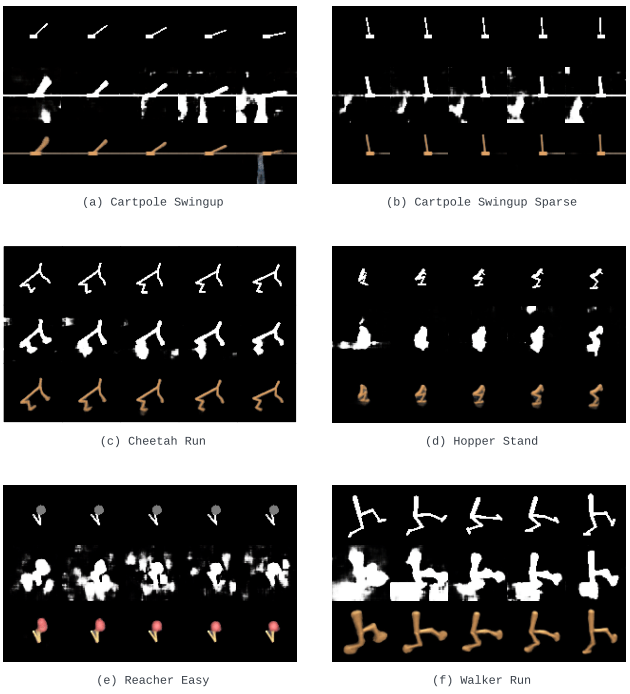


Figure 10: From the top row to the bottom row: (1) ground-truth segmentation masks, (2) SD$^{\text{approx.}}$ binary mask predictions, and (3) SD$^{\text{approx.}}$ RGB predictions.

F Distracting DMC Setup

We follow the DBC (Zhang et al., 2021) implementation to replace the background with color videos. The ground plate is also presented in the distracting environment. We used hold-out videos as background for testing. We sampled 100 videos for training from the Kinetics 400 training set of the 'driving car' class, and test-time videos were sampled from the validation set of the same class.

G Distracting Meta-World Setup

We test on six tasks from Meta-World-V2. For all tasks, we use the `corner3` camera viewpoint. The maximum episode length for Meta-World tasks is 500 environment steps, with the action repeat of 2 (making 250 policy decision steps). We classify these tasks into `easy`, `medium`, and `difficult` categories based on the training curve of DreamerPro (DreamerPro trained in the standard environments). Coffee Button, Drawer Close, and Handle Press are classified as `easy`, and we train baselines on these for 30K environment steps. Button Press Topdown (`medium`) is trained for 100K steps, and Door Open and Drawer Open (`difficult`) are trained for 1M environment steps.

H Results on Meta-World with Sparse Rewards

We also evaluate on sparse reward variations of the distracting Meta-World environments where a reward of 1 is only provided on timesteps when a *success* signal is given by the environment (e.g. objects are at their goal configuration). Rewards are 0 in all other timesteps. The maximum attainable episode reward is 250.

The sparse reward setting is more challenging because the less informative reward signal makes credit assignment more difficult for the RL agent. Fig. 11 shows that our method consistently achieves higher sample efficiency and better performance, showing promise for training agents robust to visual distractions without extensive reward engineering. In Meta-World experiments, TIA (Fu et al., 2021) is not included as it requires exhaustive hyperparameter tuning for new domains and is the lowest-performing method in DMC in general.

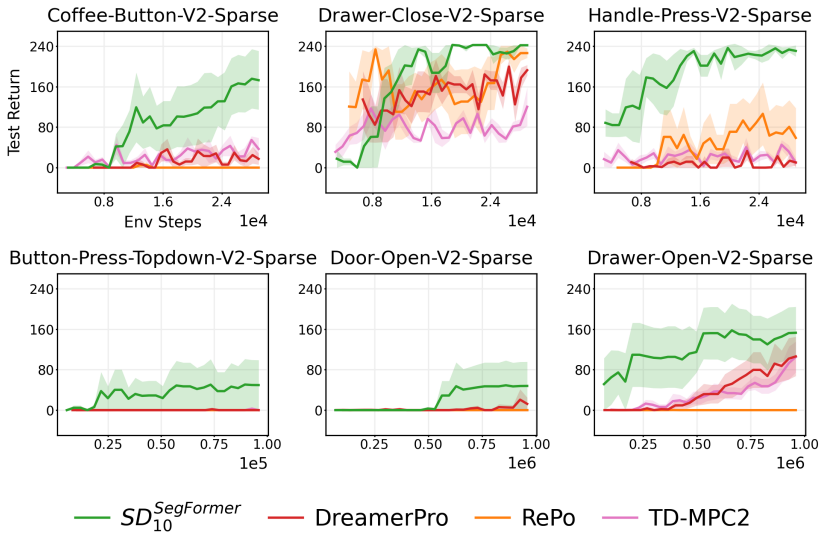


Figure 11: Learning curves on six visual robotic manipulation tasks from Meta-World with sparse rewards.

I FINE-TUNING PERSAM AND SEGFORMER

In this section, we describe how we fine-tune segmentation models and collect RGB and segmentation mask examples to adapt them.

PerSAM. Personalized SAM (PerSAM) (Zhang et al., 2023) is a segmentation model designed for personalized object segmentation building upon the Segment Anything Model (SAM) (Kirillov et al., 2023). This model is particularly a good fit for our SD use case since it can obtain a personalized segmentation model without additional training by one-shot adapting to a *single* in-domain image. In our experiments, we use the model with ViT-T as a backbone.

SegFormer. We use 5 or 10 pairs of examples to fine-tune SegFormer (Xie et al., 2021) MiT-b0.

To collect a one-shot in-domain RGB image and mask example for DMC and MetaWorld experiments, we sample a state from the initial distribution p_0 and render the RGB observation. In a few-shot scenario, we deploy a random agent in to collect more diverse observations from more diverse states.

To generate the associated masks for these states, we make additional queries to the simulation rendering API. We represent the pixel values for background and irrelevant objects as *false* and task-relevant objects as *true*. In multi-object cases, we may perform a separate adaptation operation for each task-relevant object, resulting in more than 2 mask classes. In such cases, before integrating masks with $\text{SD}^{\text{approx.}}$, we will combine the union of the mask classes for all pertinent objects as a single *true* task-relevant class, creating a binary segmentation mask compatible with our method.

In cases where example masks cannot be programmatically extracted, because such a small number of examples are required (1-10), it should also be very feasible for a human to use software to manually annotate the needed mask examples from collected RGB images.

J DETAILS ON SELECTIVE L_2 LOSS

The binary mask prediction branch in $\text{SD}^{\text{approx.}}$ is equipped with the sigmoid layer at its output. In order to obtain binary mask_{SD}, we binarize the SD binary mask prediction with a threshold of 0.9.

K DETAILS ON BASELINES

It is known that RePo (Zhu et al., 2023) outperforms many earlier works (Fu et al., 2021; Hansen et al., 2022; Zhang et al., 2021; Wang et al., 2022; Gelada et al., 2019) and that DreamerPro (Deng et al., 2022) surpasses TPC (Nguyen et al., 2021). However, theses two groups of works have been using slightly different environment setups and have not been compared with each other despite addressing the same high-level problem on the same DMC environments. In our experiments, we evaluate the representatives in each cluster on a common ground (See Appendix F) and compare them with our method.

In our experiments, we use hyperparameters used in the original papers for all the baselines, except RePo (Zhu et al., 2023) in Meta-World. RePo does not have experiments on Meta-World in which case we use hyperparameters used for Maniskill2 (Gu et al., 2023) which is another robot manipulation benchmark.

L EXTENDED RELATED WORK

There are several model-based RL approaches which introduce new auxiliary tasks. Dynalang (Lin et al., 2024) integrates language modeling as a self-supervised learning objective in world-model training. It shows impressive performance on benchmarks where the dynamics can be effectively described in natural language. However, it is not trivial to apply this method in low-level control scenarios such as locomotion control in DMC. Informed POMDP (Lambrechts et al., 2024) introduces an information decoder which uses priviledged simulator information to decode a sufficient statistic for optimal control. This shares an idea of using additional information available at training time with our method SD^{GT}. Although this can be effective on training in simulation where well-shaped

proprioceptive states exist, it cannot be applied to cases where such information is hard to obtain. In goal-conditioned RL, GAP (Nair et al., 2020) proposed to decode the difference between the future state and the goal state to help learn goal-relevant features in the state space.

M LIMITATIONS

Segmentation Dreamer achieves excellent performance across diverse tasks in the presence of distractions and provides a human interface to indicate task relevance. This capability enables practitioners to readily train an agent for their specific purposes without suffering from poor learning performance due to visual distractions. However, there are several limitations to consider.

First, since SD$^{\text{approx.}}$ harnesses a segmentation model, it can become confused when a scene contains distractor objects that resemble task-relevant objects. This challenge can be mitigated by combining our method with approaches such as InfoPower (Bharadhwaj et al., 2022), which learns controllable representations through empowerment (Mohamed & Jimenez Rezende, 2015). This integration would help distinguish controllable task-relevant objects from those with similar appearances but move without agent interaction.

Second, our method does not explicitly address randomization in the visual appearance of *task-relevant* objects, such as variations in brightness, illumination, or color. Two observations of the same internal state but with differently colored task-relevant objects may be guided toward different latent representations because our task-relevant "pixel-value" reconstruction loss forces them to be differentiated. Ideally, these observations should map to the same state abstraction since they exhibit similar behaviors in terms of the downstream task. Given that training with pixel-value perturbations on task-relevant objects is easier compared to dealing with dominating background distractors (Stone et al., 2021), our method is expected to manage such perturbations effectively without modifications. However, augmenting our approach with additional auxiliary tasks based on behavior similarity (Zhang et al., 2021) would further enhance representation learning and directly address this issue.

Finally, our approximation model faces scalability challenges when task-relevant objects constitute an open set. For instance, in autonomous driving scenarios, obstacles are task-relevant but cannot be explicitly specified. While our method serves as an effective solution when task-relevant objects are easily identifiable, complementary approaches should be considered when this assumption does not hold true.

REFERENCES

Rajaram Anantharaman, Matthew Velazquez, and Yugyung Lee. Utilizing mask r-cnn for detection and segmentation of oral diseases. In *International Conference on Bioinformatics and Biomedicine*, pp. 2197–2204. IEEE, 2018.

Marc G Bellemare, Yavar Naddaf, Joel Veness, and Michael Bowling. The arcade learning environment: An evaluation platform for general agents. *Journal of Artificial Intelligence Research*, 47: 253–279, 2013.

Homanga Bharadhwaj, Mohammad Babaeizadeh, Dumitru Erhan, and Sergey Levine. Information prioritization through empowerment in visual model-based rl. In *International Conference on Learning Representations*, 2022.

Mathilde Caron, Ishan Misra, Julien Mairal, Priya Goyal, Piotr Bojanowski, and Armand Joulin. Unsupervised learning of visual features by contrasting cluster assignments. *Advances in Neural Information Processing Systems*, 33:9912–9924, 2020.

Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. In *International Conference on Machine Learning*, pp. 1597–1607. PMLR, 2020.

Kyunghyun Cho, Bart van Merriënboer, Dzmitry Bahdanau, and Yoshua Bengio. On the properties of neural machine translation: Encoder–decoder approaches. In Dekai Wu, Marine

Carpuat, Xavier Carreras, and Eva Maria Vecchi (eds.), *Proceedings of SSST-8, Eighth Workshop on Syntax, Semantics and Structure in Statistical Translation*, pp. 103–111, Doha, Qatar, October 2014. Association for Computational Linguistics. doi: 10.3115/v1/W14-4012. URL https://aclanthology.org/W14-4012.

Fei Deng, Ingook Jang, and Sungjin Ahn. Dreamerpro: Reconstruction-free model-based reinforcement learning with prototypical representations. In *International Conference on Machine Learning*, pp. 4956–4975. PMLR, 2022.

Norm Ferns, Prakash Panangaden, and Doina Precup. Bisimulation metrics for continuous markov decision processes. *SIAM Journal on Computing*, 40(6):1662–1714, 2011.

Stefano Ferraro, Pietro Mazzaglia, Tim Verbelen, and Bart Dhoedt. Focus: Object-centric world models for robotics manipulation. *arXiv preprint arXiv:2307.02427*, 2023.

Chelsea Finn and Sergey Levine. Deep visual foresight for planning robot motion. In *International Conference on Robotics and Automation*, pp. 2786–2793. IEEE, 2017.

Xiang Fu, Ge Yang, Pulkit Agrawal, and Tommi Jaakkola. Learning task informed abstractions. In *International Conference on Machine Learning*, pp. 3480–3491. PMLR, 2021.

Carles Gelada, Saurabh Kumar, Jacob Buckman, Ofir Nachum, and Marc G Bellemare. Deepmdp: Learning continuous latent space models for representation learning. In *International Conference on Machine Learning*, pp. 2170–2179. PMLR, 2019.

Jiayuan Gu, Fanbo Xiang, Xuanlin Li, Zhan Ling, Xiqiang Liu, Tongzhou Mu, Yihe Tang, Stone Tao, Xinyue Wei, Yunchao Yao, et al. Maniskill2: A unified benchmark for generalizable manipulation skills. In *International Conference on Learning Representations*, 2023.

David Ha and Jürgen Schmidhuber. World models. *arXiv preprint arXiv:1803.10122*, 2018.

Danijar Hafner, Timothy Lillicrap, Ian Fischer, Ruben Villegas, David Ha, Honglak Lee, and James Davidson. Learning latent dynamics for planning from pixels. In *International Conference on Machine Learning*, pp. 2555–2565. PMLR, 2019.

Danijar Hafner, Timothy Lillicrap, Jimmy Ba, and Mohammad Norouzi. Dream to control: Learning behaviors by latent imagination. In *International Conference on Learning Representations*, 2020.

Danijar Hafner, Timothy Lillicrap, Mohammad Norouzi, and Jimmy Ba. Mastering atari with discrete world models. In *International Conference on Learning Representations*, 2021.

Danijar Hafner, Jurgis Pasukonis, Jimmy Ba, and Timothy Lillicrap. Mastering diverse domains through world models. *arXiv preprint arXiv:2301.04104*, 2023.

Bo Han, Quanming Yao, Xingrui Yu, Gang Niu, Miao Xu, Weihua Hu, Ivor Tsang, and Masashi Sugiyama. Co-teaching: Robust training of deep neural networks with extremely noisy labels. *Advances in Neural Information Processing Systems*, 31, 2018.

Nicklas Hansen and Xiaolong Wang. Generalization in reinforcement learning by soft data augmentation. In *International Conference on Robotics and Automation*, pp. 13611–13617. IEEE, 2021.

Nicklas Hansen, Hao Su, and Xiaolong Wang. Stabilizing deep q-learning with convnets and vision transformers under data augmentation. In *Advances in Neural Information Processing Systems*, volume 34, pp. 3680–3693, 2021.

Nicklas Hansen, Xiaolong Wang, and Hao Su. Temporal difference learning for model predictive control. In *International Conference on Machine Learning*, 2022.

Nicklas Hansen, Hao Su, and Xiaolong Wang. Td-mpc2: Scalable, robust world models for continuous control. *arXiv preprint arXiv:2310.16828*, 2023.

Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. Mask r-cnn. In *International Conference on Computer Vision*, pp. 2961–2969, 2017.

Max Jaderberg, Volodymyr Mnih, Wojciech Marian Czarnecki, Tom Schaul, Joel Z Leibo, David Silver, and Koray Kavukcuoglu. Reinforcement learning with unsupervised auxiliary tasks. In *International Conference on Learning Representations*, 2017.

Stephen James, Paul Wohlhart, Mrinal Kalakrishnan, Dmitry Kalashnikov, Alex Irpan, Julian Ibarz, Sergey Levine, Raia Hadsell, and Konstantinos Bousmalis. Sim-to-real via sim-to-sim: Data-efficient robotic grasping via randomized-to-canonical adaptation networks. In *Computer Vision and Pattern Recognition*, pp. 12627–12637, 2019.

Will Kay, Joao Carreira, Karen Simonyan, Brian Zhang, Chloe Hillier, Sudheendra Vijayanarasimhan, Fabio Viola, Tim Green, Trevor Back, Paul Natsev, et al. The kinetics human action video dataset. *arXiv preprint arXiv:1705.06950*, 2017.

Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C Berg, Wan-Yen Lo, et al. Segment anything. *arXiv preprint arXiv:2304.02643*, 2023.

Ilya Kostrikov, Denis Yarats, and Rob Fergus. Image augmentation is all you need: Regularizing deep reinforcement learning from pixels. In *International Conference on Learning Representations*, 2021.

Nathan Lambert, Brandon Amos, Omry Yadan, and Roberto Calandra. Objective mismatch in model-based reinforcement learning. In *Conference on Learning for Dynamics and Control*, 2020.

Gaspard Lambrechts, Adrien Bolland, and Damien Ernst. Informed POMDP: Leveraging additional information in model-based RL. *Reinforcement Learning Journal*, 1, 2024.

Michael Laskin, Aravind Srinivas, and Pieter Abbeel. Curl: Contrastive unsupervised representations for reinforcement learning. In *International Conference on Machine Learning*, pp. 5639–5650. PMLR, 2020.

Jessy Lin, Yuqing Du, Olivia Watkins, Danijar Hafner, Pieter Abbeel, Dan Klein, and Anca Dragan. Learning to model the world with language. In *International Conference on Machine Learning*, 2024.

Shakir Mohamed and Danilo Jimenez Rezende. Variational information maximisation for intrinsically motivated reinforcement learning. *Advances in Neural Information Processing Systems*, 28, 2015.

Suraj Nair, Silvio Savarese, and Chelsea Finn. Goal-aware prediction: Learning to model what matters. In *International Conference on Machine Learning*, pp. 7207–7219. PMLR, 2020.

Suraj Nair, Aravind Rajeswaran, Vikash Kumar, Chelsea Finn, and Abhinav Gupta. R3m: A universal visual representation for robot manipulation. In *Conference on Robot Learning*, 2022.

Tung D Nguyen, Rui Shu, Tuan Pham, Hung Bui, and Stefano Ermon. Temporal predictive coding for model-based planning in latent space. In *International Conference on Machine Learning*, pp. 8130–8139. PMLR, 2021.

Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. In *Computer Vision and Pattern Recognition*, pp. 779–788, 2016.

Julian Schrittwieser, Ioannis Antonoglou, Thomas Hubert, Karen Simonyan, Laurent Sifre, Simon Schmitt, Arthur Guez, Edward Lockhart, Demis Hassabis, Thore Graepel, et al. Mastering atari, go, chess and shogi by planning with a learned model. *Nature*, 588(7839):604–609, 2020.

Younggyo Seo, Danijar Hafner, Hao Liu, Fangchen Liu, Stephen James, Kimin Lee, and Pieter Abbeel. Masked world models for visual control. In *Conference on Robot Learning*, pp. 1332–1344. PMLR, 2022.

John So, Amber Xie, Sunggoo Jung, Jeffrey Edlund, Rohan Thakker, Ali Agha-mohammadi, Pieter Abbeel, and Stephen James. Sim-to-real via sim-to-seg: End-to-end off-road autonomous driving without real data. In *Conference on Robot Learning*, 2022.

Austin Stone, Oscar Ramirez, Kurt Konolige, and Rico Jonschkowski. The distracting control suite–a challenging benchmark for reinforcement learning from pixels. *arXiv preprint arXiv:2101.02722*, 2021.

Richard S Sutton. Dyna, an integrated architecture for learning, planning, and reacting. *ACM Sigart Bulletin*, 2(4):160–163, 1991.

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.

Emanuel Todorov, Tom Erez, and Yuval Tassa. Mujoco: A physics engine for model-based control. In *International Conference on Intelligent Robots and Systems*, pp. 5026–5033. IEEE, 2012.

Tongzhou Wang, Simon S Du, Antonio Torralba, Phillip Isola, Amy Zhang, and Yuandong Tian. Denoised mdps: Learning world models better than the world itself. In *International Conference on Machine Learning*, 2022.

Enze Xie, Wenhai Wang, Zhiding Yu, Anima Anandkumar, Jose M Alvarez, and Ping Luo. Segformer: Simple and efficient design for semantic segmentation with transformers. In *Advances in Neural Information Processing Systems*, 2021.

Denis Yarats, Rob Fergus, Alessandro Lazaric, and Lerrel Pinto. Mastering visual continuous control: Improved data-augmented reinforcement learning. *arXiv preprint arXiv:2107.09645*, 2021.

Tianhe Yu, Deirdre Quillen, Zhanpeng He, Ryan Julian, Karol Hausman, Chelsea Finn, and Sergey Levine. Meta-world: A benchmark and evaluation for multi-task and meta reinforcement learning. In *Conference on Robot Learning*, 2019.

Yuhui Yuan, Lang Huang, Jianyuan Guo, Chao Zhang, Xilin Chen, and Jingdong Wang. Ocnet: Object context network for scene parsing. *arXiv preprint arXiv:1809.00916*, 2018.

Amy Zhang, Rowan McAllister, Roberto Calandra, Yarin Gal, and Sergey Levine. Learning invariant representations for reinforcement learning without reconstruction. In *International Conference on Learning Representations*, 2021.

Marvin Zhang, Sharad Vikram, Laura Smith, Pieter Abbeel, Matthew Johnson, and Sergey Levine. Solar: Deep structured representations for model-based reinforcement learning. In *International Conference on Machine Learning*, pp. 7444–7453. PMLR, 2019.

Renrui Zhang, Zhengkai Jiang, Ziyu Guo, Shilin Yan, Junting Pan, Hao Dong, Peng Gao, and Hongsheng Li. Personalize segment anything model with one shot. *arXiv preprint arXiv:2305.03048*, 2023.

Chuning Zhu, Max Simchowitz, Siri Gadipudi, and Abhishek Gupta. Repo: Resilient model-based reinforcement learning by regularizing posterior predictability. In *Advances in Neural Information Processing Systems*, 2023.