Multi-view dreaming: multi-view world model with contrastive learning

Akira Kinose a, Masashi Okada b, Ryo Okumura b and Tadahiro Taniguchi b,c

ABSTRACT
In this paper, we propose Multi-View Dreaming, a novel reinforcement learning agent for integrated recognition and control from multi-view observations by extending Dreaming. Most current reinforcement learning methods assume a single-view observation space, and this imposes limitations on the observed data, such as lack of spatial information and occlusions. This makes obtaining ideal observational information from the environment difficult and is a bottleneck for real-world robotics applications. In this paper, we use contrastive learning to train a shared latent space between different viewpoints and show how the Products of Experts approach can be used to integrate and control the probability distributions of latent states for multiple viewpoints. We also propose Multi-View Dreaming V2, a variant of Multi-View Dreaming that uses a categorical distribution to model the latent state instead of the Gaussian distribution. Experiments show that the proposed method outperforms simple extensions of existing methods in a realistic robot control task.

CONTACT
Akira Kinose kinose.akira@jp.panasonic.com

ARTICLE HISTORY
Received 30 September 2022
Revised 23 March 2023
and 17 July 2023 Accepted
11 September 2023

KEYWORDS
World models; reinforcement learning; multimodal; robotic manipulation; sensor integration

1. Introduction
It would be desirable to have a vision-based control system that can manipulate objects even in environments with blind spots and limited image observation. In the case of a robot grasping an object on a complicated shelf, the robot must be able to control it by observing images from various cameras.

By contrast, most current reinforcement learning methods assume a single-view observation space, and this imposes limitations on the observed data, such as lack of spatial information and occlusions. This makes obtaining ideal observational data from the environment difficult, resulting in problems like missing observational data. This problem has become a bottleneck for real-world robotics applications.

Therefore, our goal in this research is to realize a method for learning control based on observations from multiple viewpoints. When solving this problem, it will be more useful for robot control in factories where multiple cameras can be installed, as well as automatic driving control where viewing information from multiple directions is required. Multi-view reinforcement learning can also be applied to research problems such as robustness to sensor degradation and multimodal data fusion. To address this problem, it is crucial to develop a model-based reinforcement learning method, which enables integrated recognition and control from multi-view observations.

Related studies, TCN [1], mfTCN [2], and MuMMI [3] address multi-view problems and are all highly relevant to our research. Compared with these, our focus is to (1) proposing “multi-view” world model which is highly needed in real-world robotics applications, and (2) systematically applying the theory to Dreaming [4] and Dreaming V2 [5] to verify its effectiveness.

In this paper, we propose Multi-View Dreaming, a model-based reinforcement learning for control based on multi-view observations. Multi-View Dreaming is a novel world model approach for integrated recognition and control from multi-view observations by extending Dreaming. Figure 1 shows an overview diagram of Multi-View Learning. We use contrastive learning to train a shared latent space between different viewpoints, and show how the Product-of-Experts [6] approach can be used to integrate and control the probability distributions of latent states for multiple viewpoints.

The proposed method is the extension of Dreaming [4] and Dreaming V2 [5] to multi-view control. Dreaming V2 focuses on representing latent states as categorical variables, while Dreaming focuses on making Dreamer decoder-free. The goal of this paper is to develop a multi-view approach to these approaches.
Figure 1. Overview of Multi-View Dreaming, the proposed world models approach. Multi-View Dreaming trains a shared latent space between different viewpoints using contrastive learning. Then, Multi-View Dreaming infers the global latent state by the Product-of-Experts for multiple latent state distributions. Using the global latent state as input observations, the agent can train a policy based on multiple viewpoint observations through reinforcement learning.

The key contributions of this paper are summarized as follows:

- **Learning world models from multi-view observations.** Using contrastive learning, the proposed Multi-View Dreaming and its categorical version variant, Multi-View Dreaming V2, train a shared latent space between different viewpoints. They then use the Product-of-Experts [6] to infer the global latent state for multiple latent state distributions.

- **Practical experiments for visual control.** We demonstrate the effectiveness of the proposed method in some scenarios, which correspond to real-world problems and realistic robot control tasks.

The remainder of this paper is organized as follows. In Section 2, key differences from related work are discussed. In Section 3, our proposed method Multi-View Dreaming / Multi-View Dreaming V2 is specified. In Section 4, the effectiveness of our proposed methods is demonstrated via simulated evaluations. Finally, in Section 5, concludes this paper.

2. Related works

World Models

Our research focused on learning world models and policies from high-dimensional observations in partially observable Markov decision process (POMDP [7]). Several approaches have been proposed to learn latent space dynamics models and use them to solve POMDP in model-based RL [8,9].

World models are a model-based reinforcement learning that uses observation data to learn a predictive model based on the agent’s behavior. World model trains a latent state dynamics model from the agent’s experience, which is used to learn behavior. It is advantageous to learn a compact state representation for high-dimensional input information such as images and to use world models to predict the future in latent space. Representative studies are the world models [10], SLAC [11], PlaNet [12], PlaNet-Bayes [13] Dreamer [14], Dreamer V2 [15], Dreaming [4], Dreaming V2 [5], etc. Learning from multiple observations is important in real-world robotics applications, but these methods do not address this issue.

The proposed method can also be seen as an extension of the world model to multi-view control, as it can infer the state representation from image observations and predict the future state of the environment in time series using a latent dynamics model. Our method is especially based on Dreaming [5].

Contrastive Learning in RL

Contrastive learning [16,17] is a self-supervised learning framework for learning useful representations by imposing similarity constraints on the latent space between training data.

In contrastive learning, the distance of image pairs in the latent space is represented as a loss function. Contrastive Learning trains image pairs with similarity constraints in the latent space so that they are close to each other if they are data augmented instances, and far from each other if they are different instances.

Works that use contrastive learning for reinforcement learning include CURL [18], Dreaming [4], CFM [19], CVRL [20], TPC [21].

Multi-View Learning in RL

Several reinforcement learning methods have been proposed to train a policy based on observed data from multiple modalities [1-3,22,23].

MuMMI [3] is a research that is particularly relevant to this paper. Both MuMMI and Multi-View Dreaming are approaches for integrating information from multiple sensors or views, but they differ in their implementation and emphasis. MuMMI is a multi-modal deep latent state-space model trained using mutual information lower-bound. It focuses on encouraging a shared representation between modalities and uses a specially
design density ratio estimator to promote consistency between the latent codes of each modality. The main focus of MuMMI is to learn useful and robust world models in the presence of missing observations. On the other hand, the advantage of Multi-View Dreaming is that it provides a flexible framework for integrating multi-view observations that can handle partial information from different views. Additionally, the use of categorical latent space in Multi-View Dreaming is a new and unexplored approach in robotics that has shown promise in improving performance.

TCN [1] and mfTCN [2] are examples of research dealing with multi-view contrastive learning, but both of them treating multi-view image embeddings as states, and using them to learn a policy. Also, 3DKeypoint [24] learns temporally-consistent 3d keypoints from multiple cameras and jointly learns a policy based on that keypoints. However, they do not sufficiently account for the fact that the environment is a POMDP and do not train a latent dynamics model. The difference between our method and these studies is that our method is model-based and can predict future states in time series, and it can integrate state representations deduced from multiple viewpoints.

In recent years, multi-viewpoint learning methods applied Neural Radiance Fields (NeRF) to reinforce learning have been proposed [25,26]. As compared to these methods, our method is an extended variant of the modality integration of the world model and inherits its extensibility and flexibility; therefore, it can work in tasks e.g. the agent or viewpoint itself is moving.

3. Multi-View dreaming

In this paper, we present Multi-View Dreaming, a model-based reinforcement learning method with world models that learns latent dynamics and a policy from multi-view observations, as an extension of Dreaming [4]. Dreaming is a model-based reinforcement learning that allows for the learning of latent state-space models without the need for reconstructing pixel observations. Dreaming is an extension of Dreamer, which is another model-based reinforcement learning.

In general, in order to address multiple observation data in a state-space model, the latent space of each modality must be commmonized and integrated. We employ contrastive learning to train a common latent space across various viewpoints and use Product-of-Experts to integrate the probabilistic distributions of the latent states across multiple viewpoints. Figure 2 shows a detailed diagram of the proposed method. In this method, we apply contrastive learning between multi-view observations to train a world model, based on the idea that images obtained from multi-view observations are augmentations of the same environment instance. Therefore, images from multi-view observations are trained to be close to each other in latent space at the same time. To recognize that observations from different viewpoints have the same latent state, the agent learns world models.

3.1. World model learning based on recurrent state-space model

The world model can learn a predictive model from the agent’s experience and use the prediction model to learn the behavior. Compact state representations are learned when trained on high-dimensional observations as images, allowing forward predictions in the learned latent space. This kind of model that predicts the future on latent space is called the latent dynamics model. By modeling the latent dynamics model of the environment, the agent can predict the long-term future and optimize its behavior without image reconstructions.

Multi-View Dreaming consists of a recurrent state-space model (RSSM) to predict forward dynamics in partially observable environments, and a reward predictor. RSSM is an important component for learning latent dynamics, and it has been used in many world models [4,14,15]. The model components are:

\[
\begin{align*}
\text{Recurrent model} & : h_t = f_\phi(h_{t-1}, z_{t-1}, a_{t-1}) \\
\text{Representation model} & : z_t \sim q_\phi(z_t \mid h_t, x_t) \\
\text{Transition predictor} & : \hat{z}_t \sim p_\phi(\hat{z}_t \mid h_t) \\
\text{Reward predictor} & : \hat{r}_t \sim p_\psi(\hat{r}_t \mid h_t, z_t) \\
\text{Actor} & : \hat{a}_t \sim p_\psi(\hat{a}_t \mid \hat{z}_t) \\
\text{Critic} & : v_\xi(\hat{z}_t) \approx E_{p_\phi,p_\psi} \left[ \sum_{\tau \geq t} \gamma^{\tau-t} \hat{r}_\tau \right] 
\end{align*}
\]

Figure 3 illustrates the detailed model architecture of Multi-View Dreaming. In the proposed model, latent states of the multiple viewpoints $z_{t1}^1, z_{t2}^2$ are integrated (details in the next section) to global stochastic latent state $z_t$.

3.2. Integration of latent state distributions

In this section, we explain how to integrate multiple latent state distributions.
Figure 2. Detailed diagram of multi-view contrastive learning. The image pairs of each viewpoint at the same time (green arrows rgb 0.439,0.678,0.278 ↔) are positive samples, and the latent space is learned so that these images are located close to each other. The image pairs of the same and different viewpoints at different times (red arrows rgb 0.752,0.223,0.211 ↔) are negative samples, and the latent space is learned so that these images are located far from each other.

Figure 3. Model architecture of Multi-View Dreaming. In this figure, assume that the model observes images from two viewpoints at the same time. The training image \( x_1^t, x_2^t \) for each viewpoint is encoded using a shared encoder. The RSSM uses a sequence of deterministic recurrent states \( h_t \). At each step, this model infers global posterior probability states \( z_t \) and prior probability states \( \hat{z}_t \). The representation model infers posterior probability states \( z_1^t \) and \( z_2^t \) for each viewpoint from current images \( x_1^t, x_2^t \) for each viewpoint and recurrence states \( h_t \). The global posterior probability state \( z_t \) is calculated from the posterior probability states \( z_1^t \) and \( z_2^t \) of each viewpoint by Product-of-Experts. The Transition predictor calculates \( \hat{z}_t \), a prior probability state that attempts to predict the posterior probability state without accessing the current image. This method uses the same decoder-free world model as Dreaming, but we train the decoder experimentally without computing the gradient to the loss function. Also, \( a_t \) represents the action and \( R_w \) represents the reward function.

**Multi-View Dreaming (Gaussian)**

The RSSM based on Dreamer assumes a Gaussian distribution for the latent state distribution. By integrating the latent states of each of the multiple viewpoints into the global latent state, the global latent state can be seen as representing the true latent state of the environment. The stochastic state \( z_t \) integrates the states of multiple viewpoints by taking a weighted harmonic mean over the mean \( \mu \) and variance \( \sigma \) of the normal distribution, as shown in the following equation:

\[
\mu_V = \frac{\sum_{v=1}^{V} \mu_v}{\sum_{v=1}^{V} 1/\sigma_v^2}, \quad \sigma_V^2 = \frac{1}{\sum_{v=1}^{V} 1/\sigma_v^2}
\]

where \( V \) denotes the number of viewpoints.

It is inspired by the Product-of-Experts [6] proposed by Hinton. The idea is to multiply the density functions of multiple probability distributions (experts) to combine them.

**Multi-View DreamingV2 (Categorical)**

In this paper, we also propose a variant of Multi-View Dreaming that uses a categorical distribution to model the latent state instead of the Gaussian distribution that was proposed in DreamerV2. There has been no prior research that uses categorical distributions for latent states of the world model to learn multi-view observations of our knowledge. In this paper, we call this method Multi-View DreamingV2. Multi-View DreamingV2 is based on DreamingV2 instead of Dreaming and is extended to multi-view observations. Because the latent state in Multi-View DreamingV2 is categorical rather than Gaussian, the Product of Experts is calculated by averaging each dimension of the categorical distribution.

Using categorical variables to represent the distribution of latent states in a system can provide greater flexibility and better capture probabilistic transitions, as opposed to assuming a unimodal Gaussian distribution.
which may be too restrictive. The distribution of multiple categorical variables could be more predictive of multimodal change.

3.3. Multi-View contrastive learning

As described in the previous section, to integrate the latent space in multiple viewpoints, learning a common world model across all viewpoints is necessary. We propose to learn a common world model for all viewpoints using contrastive learning between viewpoints for representation model.

The objective function of Multi-View Dreaming is basically the same as that of Dreaming [4]. Dreaming introduces a reconstruction-free objective derived from the ELBO objective:

$$J^\text{Dreaming} := \sum_{k=0}^{K} \left( J^\text{NCE}_k + J^\text{KL}_k \right)$$

where $K$ represents the overshooting distance. Here, we introduce the concept of overshooting and predict the future $K$ steps ahead of the overshooting distance. According to this objective function, the proposed method trains RSSM and Actor-Critic.

$J^\text{KL}_k$ is a multi-step Kullback–Leibler objective as shown below:

$$J^\text{KL}_k := \mathbb{E}_{p(z_t | z_{t-k}, a_{<t})} q(z_{t-k} | x_{t-k}, a_{<t-k}) \left[ \text{KL} \left( q(z_{t+1} | x_{t+1}, a_{<t+1}) \| p(z_{t+1} | z_t, a_t) \right) \right]$$

$J^\text{NCE}_k$ is a categorical cross-entropy objective to discriminate positive pair $(z_t, x_t)$ and negative pair $(z_t, x' \neq x_t)$ as shown below:

$$J^\text{NCE}_k := \mathbb{E}_{\tilde{p}(z_t | z_{t-k}, a_{<t})} q(z_{t-k}) \left[ \log p(z_t | x_t) - \log \sum_{x'} p(z_t | x') \right]$$

where $p(z_t | x_t)$ can be considered as a discriminator for positive pairs. Dreaming calculates $J^\text{NCE}_k$ by a random image augmentation using image cropping.

In our method, images from each viewpoint observed at the same time are selected in addition to random cropping data augmentation to increase the number of positive sample pairs. The negative samples are also sampled from images from different viewpoints observed at different times. This is based on the intuition that random cropping in contrastive learning can be regarded as equivalent to a change in viewpoint in a reinforcement learning task. Even when the viewpoint and observation image are different, we believe that the latent space representations of the same scene should be close together. Therefore, by embedding the latent states of different viewpoints at the same time in close proximity, these integrated latent states will be closer to the true latent states. We call this approach as multi-view contrastive learning.

As shown in Figure 2, the image pairs of each viewpoint at the same time are positive samples, and positive samples are image pairs taken from each viewpoint at the same time, and the latent space is learned so that these images are close to each other. Negative samples are image pairs of the same and different viewpoints at different times, and the latent space is learned so that these images are far apart.

4. Experiments

As shown Figure 4, we evaluated Multi-View Dreaming’s effectiveness in two scenarios that mimic real-world problems. We also used Robosuite [27] to demonstrate the proposed method’s effectiveness in a real-world robot control task.

**Experimental Settings**

Our main baselines are Dreamer [14], DreamerV2 [15], Dreaming [4], DreamingV2 [5], the representative model-based reinforcement learning methods. However, we did not simply compare the single-view and multi-view approaches, but also extended the baseline as follows: a simple extension of the baseline by overlaying multi-view images in the color channel direction as input images. For example, if the image of the 64 (height) × 64 (width) × 3 (color) array is observed from two viewpoints, the agent will observe the image of the 64 (height) × 64 (width) × 6 (color) array after overlaying. Multi-View Dreaming is implemented based on DreamingV2. Therefore, the elements proposed in the research up to DreamingV2 will be inherited in this method.

We experimented with the three tasks shown in Figure 4. In all tasks, observations are pixel inputs (64 × 64) only. Reacher task and Pendulum task are provided from DeepMind Control Suite [28], Lift task is provided from Robosuite [27], but environments were augmented to provide images from multiple viewpoints.

**Scenario: Blind Reacher**

In this experiment, we assume that occlusion occurs on the observed images and that the critical information required for the task is not available from a single
viewpoint. Specifically, as Figure 4 shows, viewpoint 1 blinds the right half of the screen, and viewpoint 2 blinds the left half. We handle blinds by filling them with black pixels, so the image size is unchanged. Rewards are given based on the distance between the fingertip and the goal. In this scenario, learning the policy from a single viewpoint image is insufficient to complete the task; instead, the policy must be learned by combining information from multiple viewpoints.

**Scenario: Dual View Pendulum**

In this experiment, we assume a scenario where the camera position in the environment is changed from the default and only limited information of the task is available from a single viewpoint. Viewpoint 1 observes the pendulum from directly above, while viewpoint 2 observes the pendulum from directly beside, as shown in Figure 4. Rewards are given based on the height of the ball. It is difficult to achieve any of these viewpoints with only a single camera in one direction, and it is necessary to learn the policy by combining information from two viewpoints located in different places.

**Experimental Results**

Figure 5 shows the training progress for each scenario, and the final performance scores are shown in Table 1.

In **Blind Reacher scenario**, Multi-View DreamingV2 learned policies by using the images from the two viewpoints well. However, the performance was comparable to simple extensions of Dreamer, Dreaming, and DreamingV2. Multi-View Dreaming did not perform as well as the full observation.

In **Dual View Pendulum scenario**, all methods except DreamerV2 and Multi-View DreamingV2 performed equally well in learning.

In **Robosuite Lift scenario**, Multi-View DreamingV2 outperformed all other approaches by a significant difference. Whereas the previous two experiments did not differ from baseline, Multi-View DreamingV2 was particularly effective in this task.

This suggests that it is difficult to separate the necessary information for a task like Lift, which involves complex image information and different dynamics between viewpoints, using a simple method of overlaying images in the color channel direction, and that the proposed method is effective in estimating the latent state for each viewpoint. A reconstructed image from the global latent state is shown in Figure 6. The global latent state combines and embeds both the important information of viewpoint 1 and viewpoint 2. This can be qualitatively confirmed.

According to the experimental results, MVDreamingV2 demonstrated the highest performance in the Reacher and Lift environments, but the performance decreased in the Pendulum environment. This result may...
Figure 5. Training progress of Blind Reacher (left), Dual View Pendulum (center), and Robosuite Lift (right). The $x$-axis represents total rewards per episode and the $y$-axis represents training steps.

Table 1. Comparison of Multi-View Dreaming / V2 to simple extension of conventional model-based reinforcement learning by final performance score.

|                      | MVDreaming (ours) | MVDreamingV2 (ours) | Dreamer [14] | DreamerV2 [15] | Dreaming [4] | DreamingV2 [5] |
|----------------------|-------------------|---------------------|--------------|--------------|-------------|-------------|
| Multi-View           | ✓                 | ✓                   | ✓            | ✓            | ✓           | ✓           |
| Gaussian Latent      | ✓                 | ✓                   | ✓            | ✓            | ✓           | ✓           |
| Categorical Latent   | ✓                 | ✓                   | ✓            | ✓            | ✓           | ✓           |
| Decoder-free         | ✓                 | ✓                   | ✓            | ✓            | ✓           | ✓           |
| Reacher (multi-view) | 588.6±356.0       | 936.1±95.9          | 685.0±410.7  | 42.2±73.3    | 841.3±316.7 | 860.9±285.6 |
| Pendulum (multi-view)| 812.2±130.4       | 256.5±304.5         | 801.2±110.4  | 10.8±24.5    | 831.6±126.4 | 410.4±413.99|
| Lift (single-view)   | –                 |                     | 133.5±64.28  | 132.6±62.4   | 150.5±78.5  | 330.9±113.6 |
| Lift (multi-view)    | 110.1±44.6        | 345.0±133.6         | 102.9±82.8   | 120.8±51.58  | 177.0±80.9  | 254.7±104.2 |

Figure 6. Observed image from viewpoint 1 (left). Observed image from viewpoint 2 (center). Reconstructed image from the global latent state (right).

be influenced by the fact that the Pendulum environment provides enough information to achieve the task with only single-view observations, and by the problems associated with multi-modal integration by Product-of-Experts. While multi-modal integration by Product-of-Experts can effectively aggregate information, it has been observed that incorporating experts with significantly inaccurate predictions can worsen the overall accuracy [29]. To solve this issue, it seems beneficial to adopt methods such as Mixture-of-Experts [29] or Mixture-of-Product-of-Experts [30] for modeling Multi-View.

5. Conclusion

In this paper, we proposed a world model approach for integrated recognition and control from multi-view observations by extending Dreaming. We used contrastive learning to train a shared latent space between different viewpoints and showed Product-of-Experts can integrate the probability distribution of latent states for multiple viewpoints.

We showed that the proposed method can utilize observations from multiple viewpoints in a complementary way in tasks that simulate occlusion problems and...
in realistic robot control tasks. In conclusion, Multi-View Dreaming is a world model that integrates images from different viewpoints by using Product-of-Experts and is effective for real-world problems such as occlusion. Understanding the characteristics of these methods identified in this study and making effective use of multiple viewpoints is important for the practical application of robotics. The theoretical causes of these differences are the focus of future research.

Our method can be seen as an example of a multi-view version of the generalized multimodal world model. In addition to multi-view images, world models that incorporate more modalities such as audio, tactile sensing, and depth sensors would be an intriguing one which could be usefully explored in further research. In addition, embedding domain information such as camera coordinates and robot proprioception into the latent state of each viewpoint would be a fruitful area for further work.

We were not able to experiment with the real robot in this paper, but we are currently working on a real robotics application, and evaluating it will be a future issue.

Disclosure statement
No potential conflict of interest was reported by the author(s).

Notes on contributors
Akira Kinose is currently a research engineer at Panasonic Connect Co., Ltd. He received his M.Eng. in Information Science and Engineering from Ritsumeikan University in 2020. His research involves state representation learning and large-scale language models.

Masashi Okada is currently a senior engineer at Panasonic Holdings Corp. He received his Ph.D. in Information Science from Osaka University in 2013. His research involves learning-based control methods and probabilistic machine.

Ryo Okumura is currently a senior engineer at Panasonic Holdings Corp. He received his M.Eng. in Information Science and Technology from The University of Tokyo in 2009. From 2016 to 2018, he was a visiting scholar at Stanford University in the Department of Mechanical Engineering in the School of Engineering. His research involves state representation learning and robot control.

Tadahiro Taniguchi received his Ph.D. degree from Kyoto University in 2006. From 2005 to 2008, he was a research fellow of the Japan Society for the Promotion of Science. From 2008 to 2010, he was an assistant professor at the Department of Human and Computer Intelligence, Ritsumeikan University. From 2010 to 2017, he was an associate professor in the same department. From 2015 to 2016, he was a visiting associate professor at the Department of Electrical and Electronic Engineering, Imperial College London. Since 2017, he has been a professor at the Department of Information Science and Engineering, Ritsumeikan University, and a visiting general chief scientist at Panasonic (Holdings) Corporation. He has been engaged in research on machine learning, emergent systems, and symbol emergence in robotics.

ORCID
Akira Kinose https://orcid.org/0000-0002-0714-0933
Tadahiro Taniguchi https://orcid.org/0000-0002-5682-2076

References
[1] Sermanet P, Lynch C, Chebotar Y, et al. Time-contrastive networks: self-supervised learning from video. International Conference on Robotics and Automation (ICRA), 2018.
[2] Dwibedi D, Tompson J, Lynch C, et al. Learning action-able representations from visual observations. International Conference on Intelligent Robots and Systems (IROS), 2018.
[3] Chen K, Lee Y, Soh H. Multi-modal mutual information (MUMMI) training for robust self-supervised deep reinforcement learning. In: 2021 IEEE International Conference on Robotics and Automation (ICRA). IEEE; 2021. p. 4274–4280.
[4] Okada M, Taniguchi T. Dreaming: model-based reinforcement learning by latent imagination without reconstruction. International Conference on Robotics and Automation (ICRA), 2021.
[5] Okada M, Taniguchi T. Dreamingv2: Model-based reinforcement learning with discrete world models without reconstruction. International Conference on Intelligent Robots and Systems (IROS), 2022.
[6] Hinton GE. Training products of experts by minimizing contrastive divergence. Neural Comput. 2002;14:1771–1800. doi: 10.1162/089976602760128018
[7] Kaelbling LP, Littman ML, Cassandra AR. Planning and acting in partially observable stochastic domains. Artif Intell. 1998;101(1-2):99–134. doi: 10.1016/S0004-3702(98)00023-X
[8] Zhang M, Vikram S, Smith L, et al. SOLAR: deep structured representations for model-based reinforcement learning. In: International Conference on Machine Learning (ICML), 2019.
[9] Kim T, Ahn S, Bengio Y. Variational temporal abstraction. Neural Inf Process Syst. 2019:32.
[10] Ha D, Schmidhuber J. Recurrent world models facilitate policy evolution. Neural Information Processing Systems, 2018.
[11] Lee AX, Nagabandi A, Abbeel P, et al. Stochastic latent actor-critic: deep reinforcement learning with a latent variable model. arXiv preprint arXiv:1907.00953, 2019.
[12] Hafner D, Lillicrap T, Fischer I, et al. Learning latent dynamics for planning from pixels. International Conference on Machine Learning (ICML), 2019.
[13] Okada M, Kosaka N, Taniguchi T. PlaNet of the Bayesians: reconsidering and improving deep planning network by incorporating Bayesian inference. International Conference on Intelligent Robots and Systems (IROS), 2020.
[14] Hafner D, Lillicrap T, Ba J, et al. Dream to control: learning behaviors by latent imagination. International Conference on Learning Representations (ICLR), 2020.
[15] Hafner D, Lillicrap T, Norouzi M, et al. Mastering atari with discrete world models. International Conference on Learning Representations (ICLR), 2021.

[16] Oord A.d., Li Y, Vinyals O. Representation learning with contrastive predictive coding. arXiv:1807.03748, 2018.

[17] Chen T, Kornblith S, Norouzi M, et al. A simple framework for contrastive learning of visual representations. International Conference on Learning Representations (ICLR), 2020.

[18] Srinivas A, Laskin M, Abbeel P. CURL: Contrastive unsupervised representations for reinforcement learning. International Conference on Machine Learning (ICML), 2020.

[19] Yan W, Vangipuram A, Abbeel P, et al. Learning predictive representations for deformable objects using contrastive estimation. arXiv:2003.05436, 2020.

[20] Ma X, Chen S, Hsu D, et al. Contrastive variational model-based reinforcement learning for complex observations. Conference on Robot Learning (CoRL), 2020.

[21] Nguyen TD, Shu R, Pham T, et al. Temporal predictive coding for model-based planning in latent space. International Conference on Machine Learning (ICML), 2021.

[22] Lee MA, Zhu Y, Srinivasan K, et al. Making sense of vision and touch: Self-supervised learning of multimodal representations for contact-rich tasks. International Conference on Robotics and Automation (ICRA), 2019.

[23] Li M, Wu L, Wang J, et al. Multi-view reinforcement learning. Neural Inf Process Syst. 2019;32.

[24] Chen B, Abbeel P, Pathak D. Unsupervised learning of visual 3d keypoints for control. ICML, 2021.

[25] Li Y, Li S, Sitzmann V, et al. 3d neural scene representations for visuomotor control. Conference on Robot Learning (CoRL), 2021.

[26] Driess D, Schubert I, Florence P, et al. Reinforcement learning with neural radiance fields. arXiv preprint arXiv:2206.01634, 2022.

[27] Zhu Y, Wong J, Mandlekar A, et al. robosuite: a modular simulation framework and benchmark for robot learning. arXiv:2009.12293, 2020.

[28] Tassa Y, Doron Y, Muldal A, et al. DeepMind control suite, arXiv preprint arXiv:1801.00690, 2018.

[29] Shi Y, Paige B, Torr P. Variational mixture-of-experts autoencoders for multi-modal deep generative models. Neural Inf Process Syst. 2019:32.

[30] Sutter TM, Daunhawer I, Vogt JE. Generalized multimodal ELBO. arXiv preprint arXiv:2105.02470, 2021.