

Multi-View Dreaming: Multi-View World Model with Contrastive Learning

Akira Kinose^{1*}, Masashi Okada², Ryo Okumura², Tadahiro Taniguchi^{2,3}

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 method assumes 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 DreamingV2, 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.

I. INTRODUCTION

It would be desirable to have a vision-based control system that can manipulate objects in environments where there are many blind spots and image observation is limited. 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 method assumes 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.

¹ Akira Kinose is with Innovation Center, Connected Solutions Company, Panasonic Corporation, Japan.

² Masashi Okada and Tadahiro Taniguchi are with Digital & AI Technology Center, Technology Division, Panasonic Corporation, Japan.

³ Tadahiro Taniguchi is also with Ritsumeikan University, College of Information Science and Engineering, Japan.

* kinose.akira@jp.panasonic.com

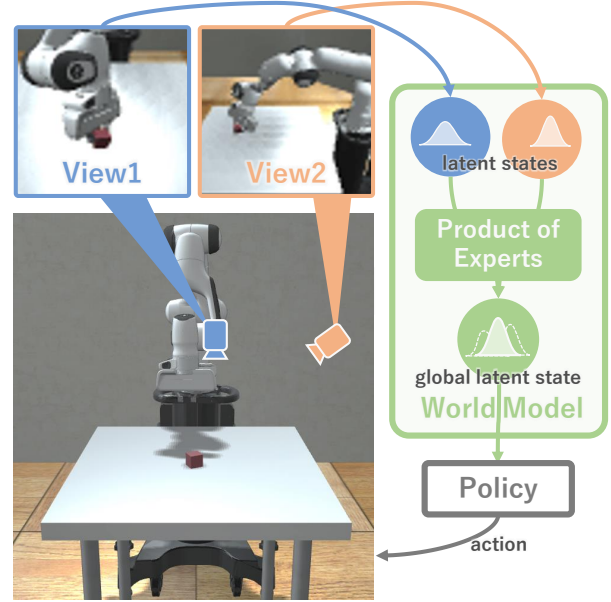


Fig. 1. Overview of Multi-View Dreaming, the proposed world models approach. Multi-View Dreaming trains a shared latent space between different viewpoints using contrastive learning. Then, Multi-View Dreaming infers the global latent state by the Product of Experts for multiple latent state distributions. By using the global latent state as input observations, the agent can train a policy based on multiple viewpoint observations through reinforcement learning.

However, model-based reinforcement learning frameworks for multi-view control have not yet been established. TCN [1] and mfTCN [2], but both of them do not consider the partial observability of the environment and do not train latent state dynamics. MuMMI [3] is multimodal reinforcement learning, which is highly relevant to this research. Compared to MuMMI, 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 DreamingV2 [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. Fig. 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 Products of Experts approach can be used to integrate and control the probability distributions of latent states for multiple viewpoints.

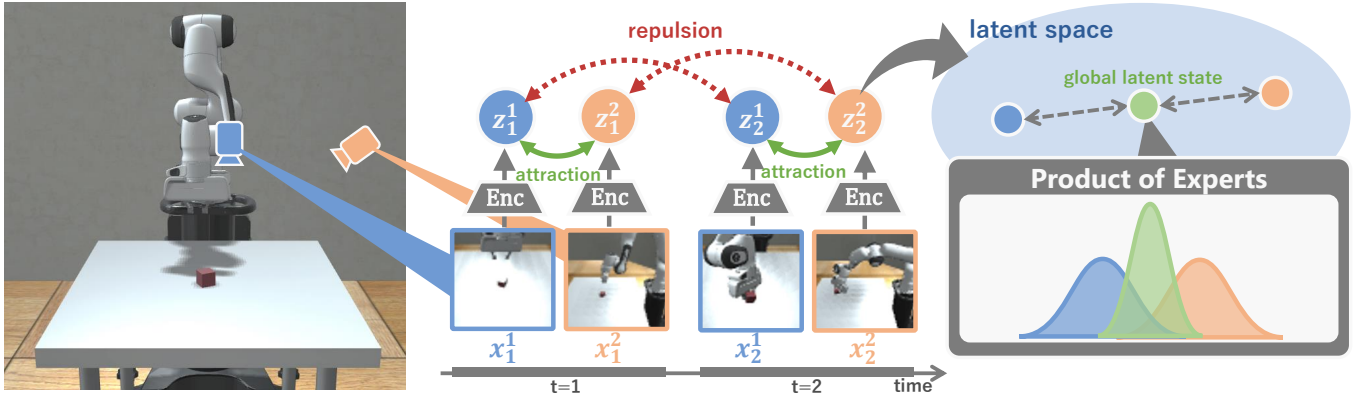


Fig. 2. Detailed diagram of multi-view contrastive learning. The image pairs of each viewpoint at the same time (green arrows \leftrightarrow) are positive samples, and the latent space is learned so that these images are located close to each other. The image pairs of the same and different viewpoints at different times (red arrows \leftrightarrow) are negative samples, and the latent space is learned so that these images are located far from each other.

The proposed method is the extension of Dreaming [4] and DreamingV2 [5] to multi-view control. DreamingV2 focuses on representing latent states as categorical variables, while Dreaming focuses on making the dreamer decoder-free. The goal of this paper is to develop a multi-view approach to these approaches.

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 DreamingV2, train a shared latent space between different viewpoints. They then use the Product of Experts 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 Sec. II, key differences from related work are discussed. In Sec. III, our proposed method Multi-View Dreaming / Multi-View DreamingV2 is specified. In Sec. IV, the effectiveness of our proposed methods is demonstrated via simulated evaluations. Finally, In Sec. V, concludes this paper.

II. RELATED WORKS

World Models

Our research focused on learning world models and policies from high-dimensional observations in partially observable Markov decision process (POMDP [6]). Several approaches have been proposed to learn latent space dynamics models and use them to solve POMDP in model-based RL [7], [8].

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 [9], SLAC [10], PlaNet [11], PlaNet-Bayes [12] Dreamer [13], DreamerV2 [14], Dreaming [4], DreamingV2 [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 [15], [16] 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 using contrastive learning for reinforcement learning include CURL [17], Dreaming [4], CFM [18], CVRL [19], TPC [20].

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], [21], [22].

MuMMI [3] is a research that is particularly relevant to this paper. In contrast to MuMMI, what is particularly important in this work is that (1) this focuses on ”multi-view”, which is highly needed in real applications to robots, and (2) systematically applies the theory to Dreamer(V2) and Dreaming(V2) to verify its effectiveness.

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. 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.

III. 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]. Fig. 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.

World Model learning based on RSSM

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], [13], [14]. The model components are:

$$\text{RSSM} \begin{cases} \text{Recurrent model:} \\ \quad h_t = f_\phi(h_{t-1}, z_{t-1}, a_{t-1}) \\ \text{Representation model:} \\ \quad z_t \sim q_\phi(z_t | h_t, x_t) \\ \text{Transition predictor:} \\ \quad \hat{z}_t \sim p_\phi(\hat{z}_t | h_t) \end{cases} \quad (1)$$

$$\begin{aligned} &\text{Reward predictor:} \\ &\hat{r}_t \sim p(\hat{r}_t | h_t, z_t) \end{aligned} \quad (2)$$

$$\text{Actor: } \hat{a}_t \sim p_\psi(\hat{a}_t | \hat{z}_t)$$

$$\text{Critic: } v_\xi(\hat{z}_t) \approx \mathbb{E}_{p_\phi, p_\psi} \left[\sum_{\tau \geq t} \gamma^{\tau-t} \hat{r}_\tau \right] \quad (3)$$

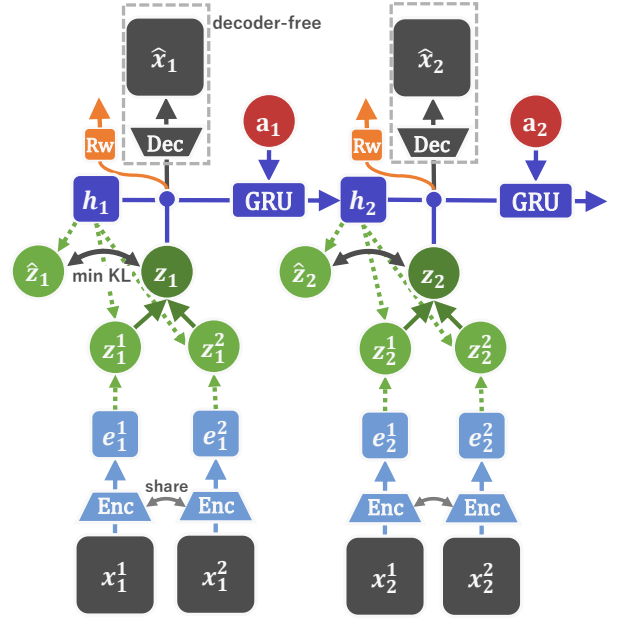


Fig. 3. Model Architecture of Multi-View Dreaming. In this figure, assume that the model observes images from two viewpoints at the same time. The training image x_t^1, x_t^2 for each viewpoint is encoded using a shared encoder. The RSSM uses a sequence of deterministic recurrent states h_t . At each 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_t^1 and z_t^2 for each viewpoint from current images x_t^1, x_t^2 for each viewpoint and recurrence states h_t . The global posterior probability state z_t is calculated from the posterior probability states z_t^1 and z_t^2 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.

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

Integration of latent state distributions

In this section, we explain how to integrate multiple latent state distributions.

1) **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 μ and variance σ of the normal distribution, as shown in the following equation:

$$\mu_V = \frac{\sum_{v=1}^V \frac{\mu_v}{\sigma_v^2}}{\sum_{v=1}^V \frac{1}{\sigma_v^2}}, \quad \sigma_V^2 = \frac{1}{\sum_{v=1}^V \frac{1}{\sigma_v^2}} \quad (4)$$

where V denotes the number of viewpoints.

It is inspired by the Products of Experts [23] proposed by Hinton. The idea is to multiply the density functions of multiple probability distributions (experts) to combine them.

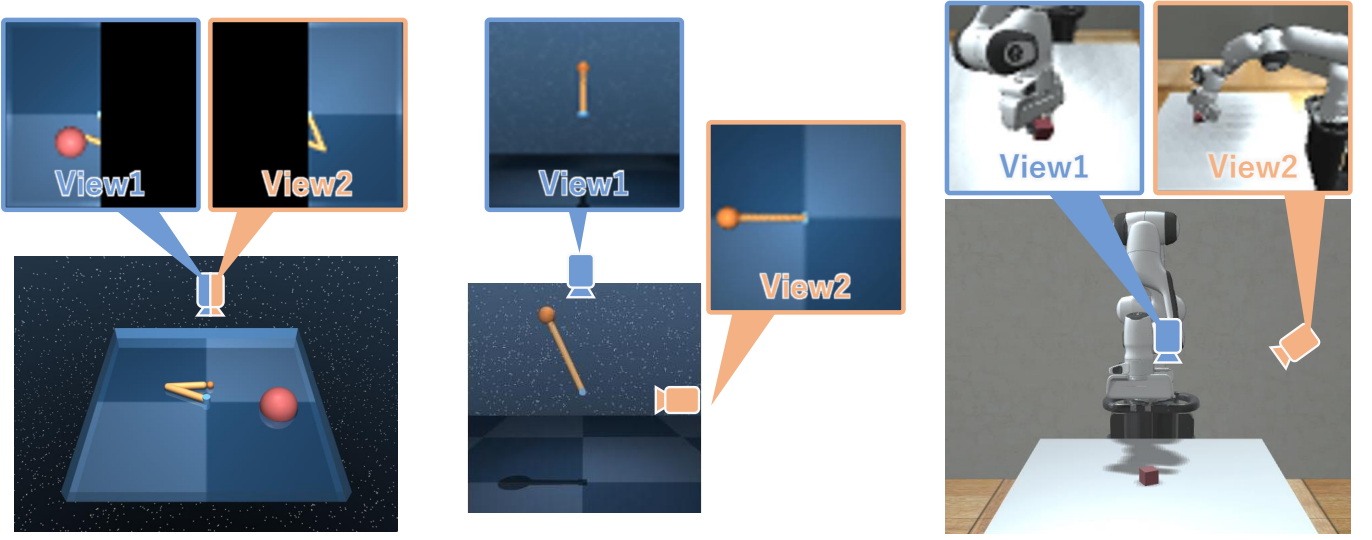


Fig. 4. Reacher-easy task uses a continuous action space with 2 dimensions (**left**). viewpoint 1 blinds the right half of the screen, and viewpoint 2 blinds the left half. Pendulum-swingup task uses a continuous action space with 1 dimensions (**center**). viewpoint 1 observes the pendulum from directly above, and viewpoint 2 observes the pendulum from directly beside. Lift-Panda task uses a continuous action space with 7 dimensions (**right**). viewpoint 1 observes the table from the front, and viewpoint 2 observes the table from the side.

In this way, the agent can choose an action based on several dimensions without covering the full dimensionality of the latent state.

2) **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 to 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.

Multi-View Contrastive Learning

As described in the previous section, to integrate the latent space in multiple viewpoints, it is necessary to learn a common world model across all viewpoint. We propose to learn a common world model for all viewpoints by 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:

$$\mathcal{J}^{\text{Dreaming}} := \sum_{k=0}^K (\mathcal{J}_k^{\text{NCE}} + \mathcal{J}_k^{\text{KL}}) \quad (5)$$

where K represents the overshooting distance. $\mathcal{J}_k^{\text{KL}}$ is a multi-step objective and $\mathcal{J}_k^{\text{NCE}}$ 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:

$$\mathcal{J}_k^{\text{NCE}} := E_{\tilde{p}(z_t | z_{t-k}, a_{<t} q(z_{t-k} \cdot))} \left[\log p(z_t | x_t) - \log \sum_{x'} p(z_t | x') \right] \quad (6)$$

Dreaming calculates $\mathcal{J}_k^{\text{NCE}}$ by 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 Fig.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.

IV. EXPERIMENTS

As shown Fig. 4, we evaluated Multi-View Dreaming's effectiveness in two scenarios that mimic real-world prob-

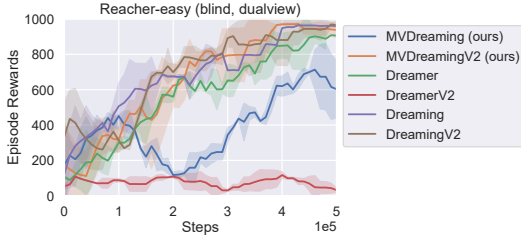


Fig. 5. Training progress of Blind Reacher scenario

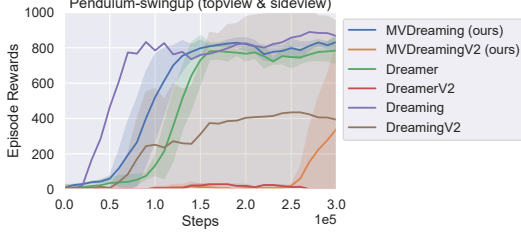


Fig. 6. Training progress of Dual View Pendulum scenario

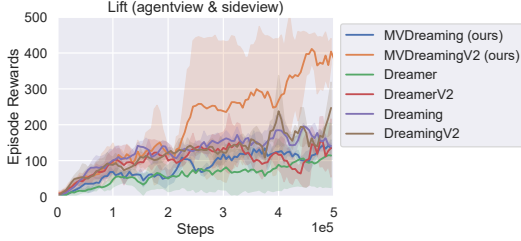


Fig. 7. Training progress of Robosuite Lift scenario

lems. We also used Robosuite to demonstrate the proposed method’s effectiveness in a real-world robot control task.

Experimental Settings

Our main baselines are Dreamer [13], DreamerV2 [14], 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 $64(\text{height}) \times 64(\text{width}) \times 3(\text{color})$ array is observed from two viewpoints, the agent will observe the image of the $64(\text{height}) \times 64(\text{width}) \times 6(\text{color})$ 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 Fig. 4. In all tasks, observations are pixel inputs (64×64) only. Reacher task and Pendulum task are provided from the DeepMind Control Suite [24], Lift task is provided from the Robosuite [25], 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,

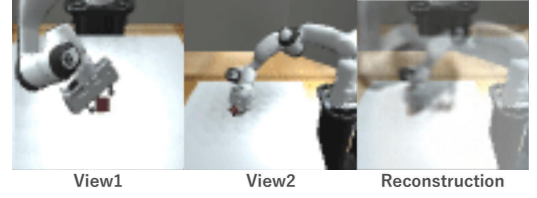


Fig. 8. Observed image from viewpoint 1 (left). Observed image from viewpoint 2 (center). Reconstructed image from the global latent state (right).

as Fig. 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. 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 Fig. 4. 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.

Scenario: Robosuite Lift

We used Robosuite in this experiment to test the effectiveness of our method in a realistic robot control task. In a realistic robot task, the robot’s body and arms act as obstacles in the observation space, necessitating multi-modal control from multiple viewpoints.

Experimental Results

Fig. 5, 6, 7 show the training progress for each scenario, and the final performance scores are shown in Table 1.

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

In the 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

TABLE I

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

	MVDreaming (ours)	MVDreamingV2 (ours)	Dreamer [13]	DreamerV2 [14]	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

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 Fig. 8. The global latent state combines and embeds both the important information of viewpoint 1 and viewpoint 2. This can be qualitatively confirmed.

V. CONCLUSION

In this paper, we proposed a novel 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 how the Products of Experts approach can be used to integrate and control the probability distributions of latent states for multiple viewpoints.

We demonstrated the effectiveness of the world model using multi-view observations in two scenarios that correspond to problems in real environments and in realistic robot control tasks. In conclusion, simple extensions of methods of overlapping images are effective for simple tasks, but a multi-view contrastive learning approach is more effective for tasks with complex images and dynamics. Understanding the features of these methods revealed in this study, as well as making effective use of multi-view, is important for practical applications in robotics. Theoretical causes of these differences will be the subject 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 real robotics application, and evaluating it will be a future issue.

REFERENCES

- [1] P. Sermanet, C. Lynch, Y. Chebotar, J. Hsu, E. Jang, S. Schaal, S. Levine, and G. Brain, “Time-contrastive networks: Self-supervised learning from video,” in *International Conference on Robotics and Automation (ICRA)*, 2018.
- [2] D. Dwibedi, J. Tompson, C. Lynch, and P. Sermanet, “Learning actionable representations from visual observations,” 2018.
- [3] K. Chen, Y. Lee, and H. Soh, “Multi-modal mutual information (MUMMI) training for robust self-supervised deep reinforcement learning,” in *2021 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 4274–4280, IEEE, 2021.
- [4] M. Okada and T. Taniguchi, “Dreaming: Model-based reinforcement learning by latent imagination without reconstruction,” in *International Conference on Robotics and Automation (ICRA)*, 2021.
- [5] M. Okada and T. Taniguchi, “Dreamingv2: Model-based reinforcement learning with discrete world models without reconstruction,” *arXiv preprint arXiv:2203.00494*, 2022.
- [6] L. P. Kaelbling, M. L. Littman, and A. R. Cassandra, “Planning and acting in partially observable stochastic domains,” *Artificial intelligence*, vol. 101, no. 1-2, pp. 99–134, 1998.
- [7] M. Zhang, S. Vikram, L. Smith, P. Abbeel, M. Johnson, and S. Levine, “SOLAR: Deep structured representations for model-based reinforcement learning,” in *International Conference on Machine Learning (ICML)*, 2019.
- [8] T. Kim, S. Ahn, and Y. Bengio, “Variational temporal abstraction,” *Neural Information Processing Systems*, 2019.
- [9] D. Ha and J. Schmidhuber, “Recurrent world models facilitate policy evolution,” in *Neural Information Processing Systems*, 2018.
- [10] A. X. Lee, A. Nagabandi, P. Abbeel, and S. Levine, “Stochastic latent actor-critic: Deep reinforcement learning with a latent variable model,” *arXiv preprint arXiv:1907.00953*, 2019.
- [11] D. Hafner, T. Lillicrap, I. Fischer, R. Villegas, D. Ha, H. Lee, and J. Davidson, “Learning latent dynamics for planning from pixels,” in *International Conference on Machine Learning (ICML)*, 2019.
- [12] M. Okada, N. Kosaka, and T. Taniguchi, “PlaNet of the Bayesians: Reconsidering and improving deep planning network by incorporating Bayesian inference,” 2020.
- [13] D. Hafner, T. Lillicrap, J. Ba, and M. Norouzi, “Dream to control: Learning behaviors by latent imagination,” *International Conference on Learning Representations (ICLR)*, 2020.
- [14] D. Hafner, T. Lillicrap, M. Norouzi, and J. Ba, “Mastering atari with discrete world models,” in *International Conference on Learning Representations (ICLR)*, 2021.
- [15] A. v. d. Oord, Y. Li, and O. Vinyals, “Representation learning with contrastive predictive coding,” *arXiv:1807.03748*, 2018.
- [16] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, “A simple framework for contrastive learning of visual representations,” in *International Conference on Learning Representations (ICLR)*, 2020.
- [17] A. Srinivas, M. Laskin, and P. Abbeel, “CURL: Contrastive unsupervised representations for reinforcement learning,” in *International Conference on Machine Learning (ICML)*, 2020.
- [18] W. Yan, A. Vangipuram, P. Abbeel, and L. Pinto, “Learning predictive representations for deformable objects using contrastive estimation,” *arXiv:2003.05436*, 2020.

- [19] X. Ma, S. Chen, D. Hsu, and W. S. Lee, “Contrastive variational model-based reinforcement learning for complex observations,” *Conference on Robot Learning (CoRL)*, 2020.
- [20] T. D. Nguyen, R. Shu, T. Pham, H. Bui, and S. Ermon, “Temporal predictive coding for model-based planning in latent space,” in *International Conference on Machine Learning (ICML)*, 2021.
- [21] M. A. Lee, Y. Zhu, K. Srinivasan, P. Shah, S. Savarese, L. Fei-Fei, A. Garg, and J. Bohg, “Making sense of vision and touch: Self-supervised learning of multimodal representations for contact-rich tasks,” in *International Conference on Robotics and Automation (ICRA)*.
- [22] M. Li, L. Wu, J. Wang, and H. Bou Ammar, “Multi-view reinforcement learning,” *Neural Information Processing Systems*, 2019.
- [23] G. E. Hinton, “Training products of experts by minimizing contrastive divergence,” *Neural computation*, 2002.
- [24] Y. Tassa, Y. Doron, A. Muldal, T. Erez, Y. Li, *et al.*, “DeepMind control suite,” *arXiv preprint arXiv:1801.00690*, 2018.
- [25] Y. Zhu, J. Wong, A. Mandlekar, and R. Martín-Martín, “robosuite: A modular simulation framework and benchmark for robot learning,” *arXiv:2009.12293*, 2020.