Causal Discovery for Modular World Models

Anonymous Author(s)
Affiliation
Address
email

Abstract

Latent world models allow agents to reason about complex environments with high-dimensional observations. However, adapting to new environments and effectively leveraging previous knowledge remain significant challenges. We present *variational causal dynamics* (VCD), a structured world model that exploits the invariance of causal mechanisms across environments to achieve fast and modular adaptation. VCD identifies reusable components across different environments by combining causal discovery and variational inference to learn a latent representation and transition model jointly in an unsupervised manner. In evaluations on simulated environments with image observations, we show that VCD is able to successfully identify causal variables. Moreover, given a small number of observations in a previously unseen, intervened environment, VCD is able to identify the sparse changes in the dynamics and to adapt efficiently. In doing so, VCD significantly extends the capabilities of the current state-of-the-art in latent world models.

1 Introduction

The ability to adapt flexibly and efficiently to novel environments is one of the most distinctive and compelling features of the human mind. It has been suggested that humans do so by learning internal models which not only contain abstract representations of the world, but also encode generalisable, structural relationships within the environment [5]. This latter aspect, it is conjectured, is what allows humans to adapt efficiently and selectively. Recent efforts have been made to mimic this kind of representation in machine learning. *World models* [e.g. 12] aim to capture the dynamics of an environment with a predictive model. Advances in latent variable models have enabled the learning of world models in a compact latent space [12, 35, 14, 9, 40] from high-dimensional observations such as images. Whilst these models have enabled agents to act in complex environments via planning [e.g. 14, 31] or learning parametric policies [e.g. 13, 12], structurally adapting to changes in the environment remains a significant challenge. The consequence of this limitation is particularly pronounced when deploying agents to environments, where distribution shifts occur. As such, we argue that it is beneficial to build structural world models that afford modular and efficient adaptation, and that *causal* modeling offers a tantalising prospect to discover such structure from observations.

Causality plays a central role in understanding distribution changes, which can be modelled as causal interventions [30]. The Sparse Mechanism Shift hypothesis [30, 6] (SMS) states that naturally occurring shifts in the data distribution can be attributed to sparse and local changes in the causal generative process. This implies that many causal mechanisms remain *invariant* across domains [29, 27, 38]. In this light, learning a *causal* model of the environment enables agents to reason about distribution shifts and to exploit the invariance of learnt causal mechanisms across different environments. Hence, we posit that world models with a causal structure can facilitate modular transfer of knowledge. To date, however, methods for causal discovery [32, 26, 28, 8, 19] require access to abstract causal variables. These are not typically available in the context of world model learning, where we wish to operate directly on high-dimensional observations.

Submitted to 36th Conference on Neural Information Processing Systems (NeurIPS 2022). Do not distribute.

In order to benefit from the structure of causal models and the ability to represent high-dimensional observations, we propose *Variational Causal Dynamics* (VCD), which combines causal discovery with variational inference. Specifically, we train a latent state-space model with a structural transition model using variational inference and sparsity regularisation from causal discovery. By jointly training a representation and a transition model, VCD learns a causally factorised world model that can modularly adapt to different environments. The key intuition behind our approach is that, since sparse causal structures can only be discovered on abstract causal variables, training the representation and the causal discovery module in an end-to-end manner acts as an inductive bias that encourages causally meaningful representations. By leveraging the learnt causal structure, VCD is able to identify the sparse mechanism changes in the environment and re-learn *only* the intervened mechanisms. This enables fast and modular adaptation to changes in dynamics.

2 Background

In a complex environment with high-dimensional observations, such as images, learning a compact latent state space dynamics model of the environment has been successful in capturing scene dynamics [14, 9]. Given a dataset of sequences $\{(o^{0:T}, a_i^{0:T})\}_{i=0}^N$, with observations o^t and actions a^t at discrete timesteps t, a generative model of the observations can be defined using latent states $z^{0:T}$ as

$$p(o^{0:T}, a^{0:T}) = \int \prod_{t=0}^{T} p_\theta(o^t|z^t)p(a^t|z^t)p_\theta(z^t|z^{t-1}, a^{t-1})dz^{0:T}, \tag{1}$$

where $p_\theta(o^t|z^t)$ and $p_\theta(z^t|z^{t-1}, a^{t-1})$ are the observation model and the transition model respectively. The variational evidence lower bound can be written as

$$\mathbb{ELBO}(\theta, \phi) = \sum_{t=0}^{T} \mathbb{E}_{q_\phi(z^t|o^t)}\big[log(p_\theta(o^t|z^t))\big] - \mathbb{E}_{q_\phi(z^{t-1}|o^{t-1})}\big[\mathbb{KL}[q_\phi(z^t|o^t)||p_\theta(z^t|z^{t-1}, a^{t-1})]\big], \tag{2}$$

where $q_\phi(z^t|o^t)$ is a learnable approximate posterior of the observations. See app.B for the derivation. Previous works have explored various functional forms of the transition distribution [e.g. 14, 35, 40]. Despite their success, these models cannot reason about changes in the environment as they cannot structurally utilise prior knowledge learnt from different environments under distribution shift.

To this end, we argue that imposing a *causal* structure on the transition model equips the learning agent with the ability to adapt to changes in a modular and efficient manner. A causal graphical model (CGM) [28] is defined as a set of random variables $\{X_1, ..., X_d\}$, their joint distribution P_X, and a directed acyclic graph (DAG), $\mathcal{G} = (X, E)$, where each edge $(i, j) \in E$ implies that X_i is a direct cause of X_j. The joint distribution admits a causal factorisation such that

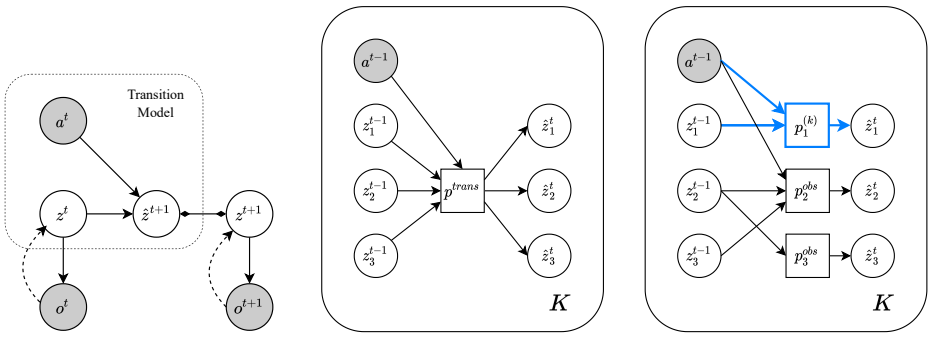
$$p(x_1, ..., x_d) = \prod_{i=0}^{d} p(x_i|PA_i), \tag{3}$$

where PA_i is the set of parents to the variable X_i in the graph. Each of the conditional distributions can be considered as an independent *causal mechanism*. In contrast to standard graphical models, CGMs support the notion of *interventions*. An intervention on the variable X_i is modelled as replacing the conditional distribution $p(x_i|PA_i)$ while leaving the other mechanisms unchanged. Given the set of intervention targets $I \subset X$, the interventional distribution can be written as

$$p'(x_1, ..., x_d) = \prod_{i \notin I} p(x_i|PA_i) \prod_{i \in I} p'(x_i|PA_i), \tag{4}$$

where $p'(\cdot|\cdot)$ is the new conditional distribution corresponding to the intervention. The SMS hypothesis [30] posits that distribution shifts tend to correspond to sparse changes in a causal model. Causal mechanisms thus tend to be *invariant* across environments [29, 27, 39]. In this light, we argue that a *causal* world model can leverage the invariance within distribution shifts for fast adaptation.

In order to learn a causal latent world model, we draw inspiration from causal discovery methods which learn causal structures from data. Specifically, We focus on method that are formulated as continuous optimisation [8, 19, 6] as these can be naturally incorporated into the variational inference framework. In particular, we follow the formulation in Differentiable Causal Discovery with Interventional data [8] (DCDI), which optimises a continuously parameterised probabilistic belief over graph structures and intervention targets. See Appendix C for further detail.



(a) Latent state space model (b) Typical transition model (c) VCD transition model

Figure 1: (a) The structure of latent state space models in one timestep. (b) The structure of latent transition models that are parameterised as fully connected neural network. (c) The causal transition model in VCD. Each variable has a separate conditional distribution p_i. Each mechanism is conditioned on a subset of the previous state. The blue lines highlight the environment-specific intervened mechanism, as opposed to the shared mechanisms corresponding to the black lines.

3 Variational Causal Dynamics

Similar to causal discovery with interventional data, variational causal dynamics (VCD) learns from action-observation sequences from an undisturbed environment, $(o_{(0)}^{0:T}, a_{(0)}^{o:T})$, and K intervened environments, $\{(o_{(k)}^{0:T}, a_{(k)}^{0:T})\}_{k=1}^{K}$. The general approach of VCD follows the latent state-space model framework (Fig.1a), where a latent representation is jointly learnt with a transition model by maximising the ELBO. In contrast to existing approaches that parameterise the transition probability $p(z^t|z^{t-1}, a^{t-1})$ as a feedforward neural network (Fig.1b), VCD learns a *causal* transition model by utilising inductive biases from causal discovery. Importantly, our approach is grounded in the hypothesis that causal structures can only be discovered on semantically meaningful causal representations of the system. We therefore argue that training a representation and a transition model jointly to optimise a causal discovery objective can lead to a latent representation that affords causal transition models and is hence semantically meaningful.[1] Taking the view that shifts in the dynamics can be attributed to sparse causal interventions, we posit that a causal transition model facilitates modular adaptation to new environments by leveraging the invariance of causal mechanisms.

3.1 Causal transition model

In order to discovery causal structure on transition dynamics, the transition model in VCD is designed to mimic the structure of a CGM (Eq. 4). The following is the transition model for environment k:

$$p^k(z^t|z^{t-1}, a^{t-1}) = \prod_i^d p_i^{(0)}(z_i^t|M_i^{\mathcal{G}} \odot [z^{t-1}, a^{t-1}])^{1-R_{ki}^{\mathcal{I}}} p_i^{(k)}(z_i^t|M_i^{\mathcal{G}} \odot [z^{t-1}, a^{t-1}])^{R_{ki}^{\mathcal{I}}}, \quad (5)$$

where $p^{(0)}$ is the transition model shared across all environments, $p^{(k)}$ is the interventional distribution specific to environment k, $M_i^{\mathcal{G}}$ is the binary causal masks that selects the causal parents for variable i, $R_{ki}^{\mathcal{I}}$ is a binary variable that is 1 when the variable i is an intervention target in environment k. This formulation of the transition is designed around three main architectural features (Fig. 1c):

- **independent mechanisms**. In contrast to fully connected models, the transition distribution for each latent dimension is parameterised by separate networks. This is motivated by the Independent Causal Mechanisms principle which posits that the causal generative process of the variables is composed of autonomous mechanisms [30]. In this work, each conditional distribution is parameterised as recurrent GRUs[10, 14].
- **sparse causal dependencies.** Following the structure of a CGM, we condition each variable only on its causal parents according to the learnable causal graph $\mathcal{G}$. This is modelled using the binary masks $M_i^{\mathcal{G}}$ which mask away non-parent nodes.

[1]The reader is referred to [22] for a discussion of the identifiability of causal representations in this setting.

- **Sparse interventions.** The intervention mask variables $R^{\mathcal{I}}_{ik}$ are introduced to model interventions. These variables act as switches between reusing a shared observational model, $p^{(0)}$, and an environment-specific interventional model, $p^{(k)}$. This is analogous to $p(x_i|PA_i)$ and $p'(x_i|PA_i)$ in Eq. 4.

3.2 Training

Similar to DCDI, the causal graph $\mathcal{G}$ and the intervention targets $\mathcal{I}$ are jointly trained with the model parameters. We parameterise the belief over the causal adjacency matrix $M^{\mathcal{G}}$ as a random binary matrix. Each entry $M^{\mathcal{G}}_{ij}$ follows a Bernoulli distribution with success probability $\sigma(\alpha_{ij})$, where α_{ij} is a scalar parameter and $\sigma(\cdot)$ is the sigmoid function. Similarly, a random binary matrix $R^{\mathcal{I}}$ is parameterised using the scalar variable β_{ki} for each entry. Unlike DCDI, due to the existence of latent variables, we cannot directly maximise the data likelihood. Instead, we maximise the expected ELBO across all environments over causal graphs and intervention masks with sparsity regularisation,

$$L(\theta, \phi, \alpha, \beta) = \sum_{k \in [0,1,\ldots,K]} \mathbb{E}_{\mathcal{G},\mathcal{I}} \big[\mathbb{ELBO}(o^{0:T}_{(k)}, a^{0:T}_{(k)}; \theta, \phi, \mathcal{G}, \mathcal{I}) - \lambda_G |\mathcal{G}| - \lambda_I |\mathcal{I}| \big], \tag{6}$$

where the ELBO term is given by the expression in Equation (2), in which the transition model, $p^{(k)}_\theta(z^t|z^{t-1}, a^{t-1})$, is further factorised as in Equation (5). The gradients through the expectation terms are estimated using reparameterisation tricks [18, 20].[2] Crucially, by applying sparsity regularisation on the causal graph and the intervention targets, the model is encourage to learn representation which afford sparse dependencies and sparse changes across environments.

3.3 Adaptation

Due to the modular nature of the transition model, VCD can adapt to new, unseen environments by jointly inferring the intervention targets and the new model parameters for the intervened mechanisms. One way to implement modular transfer is to train VCD on trajectories in the new environment while fixing the trained parameters for the causal graph and the mechanisms in the undisturbed environment. The new parameters can be jointly trained in an analogous way to Eq. 6. Under the SMS hypothesis, only a small subset of the mechanisms need to be adapted in a new environment. As such, by leveraging past experience, VCD can adapt more quickly to environment change.

4 Experiments

In this section, we demonstrate that VCD is able to learn from multiple environments by learning a causal world model that explicitly captures the changes between the environments as interventions. We show that VCD is able to learn representation that reflects the ground-truth causal variables of the environments, and illustrate that, in an unseen environment, VCD can leverage past experience to perform modular adaptation, resulting in significantly improved data efficiency over the baselines.

Setting. We evaluate VCD on a simulated dataset of a 2-D multi-body system which contains four particles that affect each other via a spring or an electrostatic-like force. We consider changes such as strengthening or removing one of the springs, changing the mass of a particle, or constraining the position of a particle along the x or the y axis, which can be considered as interventions on one or more causal variables. The observation is given by rendering the environment as an image.

We compare the performance of VCD against RSSM [14], a state-of-the-art latent world model. As RSSM does not support learning from multiple environments, we consider two adaptations of RSSM with different levels of knowledge transfer between environments: (1) **RSSM**, where one transition model is trained over all environments, i.e. maximum transfer of knowledge; and (2) **MultiRSSM**, where individual transition models are trained on each environment, i.e. no transfer of knowledge. We hypothesise that, compared to these two extremes of knowledge sharing, VCD is able to capture environment-specific behaviours whilst reusing invariant mechanisms via modular transfer.

Representation. One of the key hypotheses of this work is that jointly learning a representation and a transition model using causal discovery leads to causally meaningful representations. Here we examine the quality of the learnt latent space. Fig. 2 shows the image reconstructions of points

[2] For further implementation details, derivation of the lower bound, see Appendices B and C.

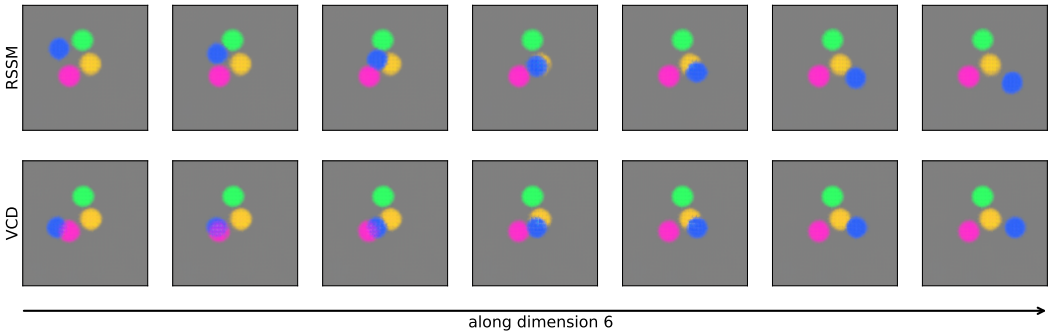


Figure 2: Reconstructed images from samples along dimension 6 of the latent space. Compared to RSSM, VCD learns an axis-aligned representation where the blue particle moves horizontally.

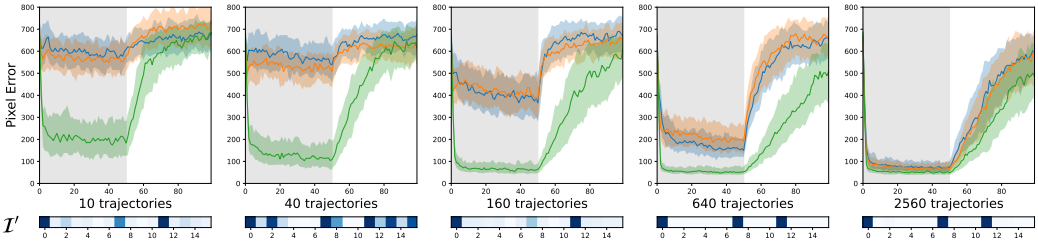


Figure 3: Rollout errors experiments where the models are trained on datasets of varying sizes (lower is better). The models receive observations for the first half of the trajectory (shaded), and perform latent space rollout for the rest. VCD significantly outperforms the other models with little data. The bar below each plot shows the learnt intervention targets in the new environment. VCD reuses most of the previously learnt mechanisms, as indicated by the sparsity in the learnt intervention targets.

drawn from a straight line along dimension 6 of the latent space. Since both models are initialised with the same encoder, this provides a qualitative intuition as to how the causal discovery inductive bias shapes the latent space. Compared to RSSM, VCD is able to learn an *axis-aligned* coordinate (x coordinate) of the blue particle. Note that the motion of bouncing off the boundaries is only separable in the x, y frame, implying that the dynamics in axis-aligned coordinates is sparser. Visualisation of the other latent dimensions is available in App. D.

Adaptation. We provide empirical evidence that VCD can adapt to a new environment with less data compared to the baselines by reusing learnt mechanisms in a modular fashion. We collect datasets of different sizes in a previously unseen intervened environment where particle 1 is constrained horizontally. RSSM and MultiRSSM adapt to the new environment by optimizing the ELBO, with the difference that MultiRSSM instantiates a new transition model randomly and RSSM initialises the transition model using pre-trained parameters. VCD performs adaptation as described in section3.3. Fig. 3 shows the rollout error of the models trained on datasets of varying size. Across all models, performance improves as the dataset grows. However, in contrast to RSSM and MultiRSSM, which overfits the dataset when the number of trajectories is small, VCD is able to predict significantly more accurately. This is because VCD estimates the intervention targets and reuses trained modules that remain invariant.

5 Conclusion and discussion

In this paper, we propose VCD, a predictive world model with a causal structure that is able to consume high-dimensional observations. This is achieved by jointly training a representation and a causally structured transition model using a modified causal discovery objective. In doing so, VCD is able to identify causally meaningful representations of the observations and discover sparse relationships in the dynamics of the system. By leveraging the invariance of causal mechanisms, VCD is able to adapt to new environments efficiently by identifying relevant mechanism changes and updating in a modular way, resulting in significantly improved data efficiency.

References

[1] K. Ahuja, J. Hartford, and Y. Bengio. Properties from mechanisms: An equivariance perspective on identifiable representation learning. *arXiv preprint arXiv:2110.15796*, 2021.

[2] M. Arjovsky, L. Bottou, I. Gulrajani, and D. Lopez-Paz. Invariant risk minimization. *arXiv preprint arXiv:1907.02893*, 2019.

[3] Rim Assouel, Lluis Castrejon, Aaron Courville, Nicolas Ballas, and Yoshua Bengio. Vim: Variational independent modules for video prediction. In *Conference on Causal Learning and Reasoning*, pages 70–89. PMLR, 2022.

[4] E. Banijamali, R. Shu, H. Bui, A. Ghodsi, et al. Robust locally-linear controllable embedding. In *International Conference on Artificial Intelligence and Statistics*, pages 1751–1759. PMLR, 2018.

[5] T. Behrens, T. Muller, J. Whittington, S. Mark, A. Baram, K. Stachenfeld, and Z. Kurth-Nelson. What is a cognitive map? organizing knowledge for flexible behavior. *Neuron*, 100(2):490–509, 2018. ISSN 0896-6273. doi: https://doi.org/10.1016/j.neuron.2018.10.002. URL https://www.sciencedirect.com/science/article/pii/S0896627318308560.

[6] Y. Bengio, T. Deleu, N. Rahaman, R. Ke, S. Lachapelle, O. Bilaniuk, A. Goyal, and C. Pal. A meta-transfer objective for learning to disentangle causal mechanisms, 2019. URL arXiv:1901.10912.

[7] I. Bica, D. Jarrett, and M. van der Schaar. Invariant causal imitation learning for generalizable policies. *Advances in Neural Information Processing Systems*, 34, 2021.

[8] P. Brouillard, S. Lachapelle, A. Lacoste, S. Lacoste-Julien, and A. Drouin. Differentiable causal discovery from interventional data. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 21865–21877. Curran Associates, Inc., 2020. URL https://proceedings.neurips.cc/paper/2020/file/f8b7aa3a0d349d9562b424160ad18612-Paper.pdf.

[9] L. Buesing, T. Weber, S. Racanière, S. M. Ali Eslami, D. Rezende, D. Reichert, F. Viola, F. Besse, K. Gregor, D. Hassabis, and D. Wierstra. Learning and querying fast generative models for reinforcement learning. *ArXiv*, abs/1802.03006, 2018.

[10] K. Cho, B. van Merrienboer, C. Gulcehre, F. Bougares, H. Schwenk, and Y. Bengio. Learning phrase representations using rnn encoder-decoder for statistical machine translation. In *Conference on Empirical Methods in Natural Language Processing (EMNLP 2014)*, 2014.

[11] Emily Denton and Rob Fergus. Stochastic video generation with a learned prior. In *International conference on machine learning*, pages 1174–1183. PMLR, 2018.

[12] D. Ha and J. Schmidhuber. Recurrent world models facilitate policy evolution. In *Advances in Neural Information Processing Systems 31*, pages 2451–2463. Curran Associates, Inc., 2018. URL https://papers.nips.cc/paper/7512-recurrent-world-models-facilitate-policy-evolution. https://worldmodels.github.io.

[13] D. Hafner, T. Lillicrap, J. Ba, and M. Norouzi. Dream to control: Learning behaviors by latent imagination. *arXiv preprint arXiv:1912.01603*, 2019.

[14] D. Hafner, T. Lillicrap, T. Fischer, T. Villegas, D. Ha, H. Lee, and J. Davidson. Learning latent dynamics for planning from pixels. In *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*. PMLR, 09–15 Jun 2019. URL https://proceedings.mlr.press/v97/hafner19a.html.

[15] A. Hauser and P. Bühlmann. Characterization and greedy learning of interventional Markov equivalence classes of directed acyclic graphs. *Journal of Machine Learning Research*, 13:2409–2464, 2012.

[16] Biwei Huang, Fan Feng, Chaochao Lu, Sara Magliacane, and Kun Zhang. AdaRL: What, where, and how to adapt in transfer reinforcement learning. In *International Conference on Learning Representations*, 2022. URL https://openreview.net/forum?id=8H5bpVwvt5.

[17] Aapo Hyvarinen, Hiroaki Sasaki, and Richard Turner. Nonlinear ica using auxiliary variables and generalized contrastive learning. In *The 22nd International Conference on Artificial Intelligence and Statistics*, pages 859–868. PMLR, 2019.

[18] E. Jang, S. Gu, and B. Poole. Categorical reparameterization with gumbel-softmax. In *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings*, 2017. URL https://openreview.net/forum?id=rkE3y85ee.

[19] R. Ke, O. Bilaniuk, A. Goyal, S. Bauer, H. Larochelle, B. Schölkopf, M. C. Mozer, C. Pal, and Y. Bengio. Learning neural causal models from unknown interventions, 2020. arXiv:1910.01075v2.

[20] D. Kingma and M. Welling. Auto-encoding variational bayes. In *2nd International Conference on Learning Representations, ICLR*, 2014. URL http://arxiv.org/abs/1312.6114.

[21] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[22] S. Lachapelle, P. López, Y. Sharma, K. Everett, R. Le Priol, A. Lacoste, and S. Lacoste-Julien. Disentanglement via mechanism sparsity regularization: A new principle for nonlinear ica, 2021. URL https://arxiv.org/abs/2107.10098.

[23] Minne Li, Mengyue Yang, Furui Liu, Xu Chen, Zhitang Chen, and Jun Wang. Causal world models by unsupervised deconfounding of physical dynamics. *arXiv preprint arXiv:2012.14228*, 2020.

[24] Yunzhu Li, Antonio Torralba, Anima Anandkumar, Dieter Fox, and Animesh Garg. Causal discovery in physical systems from videos. *Advances in Neural Information Processing Systems*, 33:9180–9192, 2020.

[25] Phillip Lippe, Sara Magliacane, Sindy Löwe, Yuki M Asano, Taco Cohen, and Stratis Gavves. Citris: Causal identifiability from temporal intervened sequences. In *International Conference on Machine Learning*, pages 13557–13603. PMLR, 2022.

[26] J. Pearl. *Causality: Models, Reasoning, and Inference*. Cambridge University Press, New York, NY, 2nd edition, 2009.

[27] J. Peters, P. Bühlmann, and N. Meinshausen. Causal inference using invariant prediction: identification and confidence intervals. *Journal of the Royal Statistical Society, Series B: Statistical Methodology (with discussion)*, 78(5):947–1012, 2016.

[28] J. Peters, D. Janzing, and B. Schölkopf. *Elements of Causal Inference - Foundations and Learning Algorithms*. MIT Press, Cambridge, MA, USA, 2017.

[29] B. Schölkopf, D. Janzing, J. Peters, E. Sgouritsa, K. Zhang, and J. Mooij. On causal and anticausal learning. In J. Langford and J. Pineau, editors, *Proceedings of the 29th International Conference on Machine Learning (ICML)*, pages 1255–1262, New York, NY, USA, 2012. Omnipress. URL http://icml.cc/2012/papers/625.pdf.

[30] B. Schölkopf, F. Locatello, S. Bauer, N. R. Ke, N. Kalchbrenner, A. Goyal, and Y. Bengio. Toward causal representation learning. *Proceedings of the IEEE - Advances in Machine Learning and Deep Neural Networks*, 109(5):612–634, 2021. doi: 10.1109/JPROC.2021.3058954. URL https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=9363924.

[31] R. Sekar, O. Rybkin, K. Daniilidis, P. Abbeel, D. Hafner, and D. Pathak. Planning to explore via self-supervised world models. In *ICML*, 2020.

[32] P. Spirtes, C. Glymour, and R. Scheines. *Causation, Prediction, and Search*. MIT Press, Cambridge, MA, 2nd edition, 2000.

[33] J. Tian and J. Pearl. Causal discovery from changes. In *Proceedings of the 17th Annual Conference on Uncertainty in Artificial Intelligence (UAI)*, pages 512–522, 2001.

[34] Ruben Villegas, Arkanath Pathak, Harini Kannan, Dumitru Erhan, Quoc V Le, and Honglak Lee. High fidelity video prediction with large stochastic recurrent neural networks. *Advances in Neural Information Processing Systems*, 32, 2019.

[35] M. Watter, J. Springenberg, J. Boedecker, and M. Riedmiller. Embed to control: A locally linear latent dynamics model for control from raw images. In C. Cortes, N. Lawrence, D. Lee, M. Sugiyama, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 28. Curran Associates, Inc., 2015. URL https://proceedings.neurips.cc/paper/2015/file/a1afc58c6ca9540d057299ec3016d726-Paper.pdf.

[36] Weiran Yao, Yuewen Sun, Alex Ho, Changyin Sun, and Kun Zhang. Learning temporally causal latent processes from general temporal data. In *International Conference on Learning Representations*, 2022. URL https://openreview.net/forum?id=RDlLMjLJXdq.

[37] A. Zhang, C. Lyle, S. Sodhani, A. Filos, M. Kwiatkowska, J. Pineau, Y. Gal, and D. Precup. Invariant causal prediction for block mdps. In *International Conference on Machine Learning*, pages 11214–11224. PMLR, 2020.

[38] K. Zhang, M. Gong, and B. Schölkopf. Multi-source domain adaptation: A causal view. In *Proceedings of the 29th AAAI Conference on Artificial Intelligence*, pages 3150–3157, 2015.

[39] K. Zhang, B. Huang, J. Zhang, C. Glymour, and B. Schölkopf. Causal discovery from nonstationary/heterogeneous data: Skeleton estimation and orientation determination. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence (IJCAI 2017)*, pages 1347–1353, 2017.

[40] 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*, pages 7444–7453. PMLR, 2019.

A Additional Related Works

VCD learns causal latent state-space models by combining variational inference and causal discovery. This section discusses related works in latent space world model learning, causal representation learning and causal discovery in relation to our work.

Predictive models of the environment can be used to derive exploration- [31] or reward-driven [12, 13, 14] behaviours. World models [12] train a representation encoder and a RNN-based transition model in a two-stage process. Other approaches [14, 40, 35] learn a generative model by jointly training the representation and the transition via variational inference. PlaNet [14] parameterises the transition model with RNNs. E2C [35, 4] and SOLAR [40] use locally-linear transition models, arguing that including constraints in the dynamics model yields structured latent spaces that are suitable for control problems. In a similar vein, the use of latent prediction models have also been explored in the context of video prediction [34, 11, 3]. Our proposed approach shares the general principle that latent representations can be shaped by structured transition mechanisms [1]. However, to the best of our knowledge, VCD is the first approach that implements a causal transition model given high-dimensional inputs in this context.

Causal discovery methods enable the learning of causal structure from data. Approaches can be categorised as constraint-based (e.g. [32]) and score-based (e.g. [15]). The reader is referred to [28] for a detailed review of causal discovery methods. Motivated by the fact that these methods require access to abstract causal variables, recent efforts have been made to reconcile machine learning, which has the ability to operate on low-level data, and causality [30]. Recent advances in this area include theoretical works on independent component analysis exploring the conditions under which disentanglement is possible [36, 17, 22]. Our current work situates within this broader context of causal representation learning, and aims to identify causally meaningful representations via the discovery of causal transition dynamics. To this end, our work builds on the theory developed in [22], where disentanglement is achieved via mechanism sparsity, by focusing on the adaptation capabilities of learnt causal models and empirically showing that the method is applicable to image observations.

There has been significant interest in leveraging causal reasoning in world model learning [23, 24]. These approaches typically rely on predefined representations such as keypoints [24] or object slots [23], and attribute changes in the dynamics to unobserved confounders, which can be estimated from observations. Our approach differs in that we couple the learning of the representation and the model, which enables dynamics-aware discovery of causal representations. Similar approaches of jointly learning representation and model structure have also been explored in different contexts such as reinforcement learning [16] and temporally intervened sequences [25]. Specifically, AdaRL [16] considers few-shot adaptation under sparse mechanism shift (SMS) by attributing changes in the environments to low-dimensional environment-specific parameters. In contrast, by explicitly drawing from causal discovery, VCD models the dynamics of the scene as independent causal mechanisms. This, in turn, allows our approach to directly identify localised mechanism shifts between environments in an interpretable manner.

Another branch of causality-inspired work leverages the invariance of causal mechanisms by learning invariant predictors across environments [33, 29, 27, 39, 2]. This invariance has been studied in the context of state abstractions in MDPs [37], and invariant policies can be learnt via imitation learning from different environments [7]. In contrast, our approach models the full generative process of the data across different environments rather than learning discriminative predictors.

B Derivations

Using the approximate posterior $q(z^{0:T}|o^{0:T}, a^{0:T}) = \prod_t q(z^t|o^t)$, the variational log lower bound for the latent state-space model (Eq. 2) can be derived from importance weighting and Jensen's inequality:

$$\log p(o^{0:T}, a^{0:T}) = \log \int \prod_{t=0}^{T} p(o^t|z^t)p(a^t|z^t)p(z^t|z^{t-1}, a^{t-1})dz^{0:T} \tag{7}$$

$$= \log \int \prod_{0}^{T} p(o^t|z^t)p(a^t|z^t)\frac{p(z^t|z^{t-1}, a^{t-1})}{q(z^t|o^t)}q(z^t|o^t)dz^{0:T} \tag{8}$$

$$\geq \mathbb{E}_{\prod q(z^t|o^t)}\left[\log\left(\prod_{o}^{T} p(o^t|z^t)p(a^t|z^t)\frac{p(z^t|z^{t-1}, a^{t-1})}{q(z^t|o^t)}\right)\right] \tag{9}$$

$$= \sum_{0}^{T} \mathbb{E}_{q(z^t|o^t)}\left[\log p(o^t|z^t) + \log p(a^t|z^t)\right] + \mathbb{E}_{q(z^{t-1}|o^{t-1})}\left[\mathbb{E}_{q(z^t|o^t)}\left[\log p(z^t|z^{t-1}, a^{t-1}) - \log p(z^t|o^t)\right]\right] \tag{10}$$

$$= \sum_{0}^{T} \mathbb{E}_{q(z^t|o^t)}\left[\log p(o^t|z^t) + \log p(a^t|z^t)\right] - \mathbb{E}_{q(z^{t-1}|o^{t-1})}\left[\mathbb{KL}\left[q(z^t|o^t)||p(z^t|z^{t-1}, a^{t-1})\right]\right]. \tag{11}$$

Since the policy $p(a^t|z^t)$ is constant with respect to the model parameters, we omit this term and write the ELBO objective as

$$\mathbb{ELBO}(\theta, \phi) = \sum_{t=0}^{T} \mathbb{E}_{q_\phi(z^t|o^t)}\left[log(p_\theta(o^t|z^t))\right] - \mathbb{E}_{q_\phi(z^{t-1}|o^{t-1})}\left[\mathbb{KL}[q_\phi(z^t|o^t)||p_\theta(z^t|z^{t-1}, a^{t-1})]\right], \tag{12}$$

where θ is the model parameter and ϕ is the parameter for the approximate posterior. θ and ϕ are omitted henceforth to simplify notation. In VCD, the KL divergence term can be further decomposed by exploiting the structure of the transition model (Eq. 5), and the assumption that variables within each timestep are independent:

$$\mathbb{KL}\left[q(z^t|o^t)||p^{(k)}(z^t|z^{t-1}, a^{t-1})\right] \tag{13}$$

$$= -\int \log\left(\frac{p^{(k)}(z^t|z^{t-1}, a^{t-1})}{q(z^t|o^t)}\right) q(z^t|o^t)dz^t \tag{14}$$

$$= -\sum_{i=0}^{d} \int \log\left(\frac{p_i^{(0)}(z_i^t|M_i^{\mathcal{G}} \odot [z^{t-1}, a^{t-1}])^{1-R_{ki}^{\mathcal{I}}} p_i^{(k)}(z_i^t|M_i^{\mathcal{G}} \odot [z^{t-1}, a^{t-1}])^{R_{ki}^{\mathcal{I}}}}{q(z_i^t|o^t)}\right) q(z_i^t|o^t)dz_i^t \tag{15}$$

$$= -\sum_{0}^{d}\left((1 - R_{ki}^{\mathcal{I}}) \int \log\left(\frac{p_i^{(0)}(z_i^t|M_i^{\mathcal{G}} \odot [z^{t-1}, a^{t-1}]}{q(z_i^t|o^t)}\right) q(z_i^t|o^t)dz_i^t + R_{ki}^{\mathcal{I}} \int \log\left(\frac{p_i^{(k)}(z_i^t|M_i^{\mathcal{G}} \odot [z^{t-1}, a^{t-1}])}{q(z_i^t|o^t)}\right) q(z_i^t|o^t)dz_i^t\right) \tag{16}$$

$$= -\sum_{0}^{d}\left((1 - R_{ki}^{\mathcal{I}})\mathbb{KL}\left[q(z_i^t|o^t)||p_i^{(0)}(z_i^t|M_i^{\mathcal{G}} \odot [z^{t-1}, a^{t-1}])\right] + R_{ki}^{\mathcal{I}}\mathbb{KL}\left[q(z_i^t|o^t)||p_i^{(k)}(z_i^t|M_i^{\mathcal{G}} \odot [z^{t-1}, a^{t-1}])\right]\right). \tag{17}$$

The KL terms can be computed analytically since the conditional distributions in the last expression are univariate Gaussian distributions. In training time, the gradients through the expectation terms in the ELBO is estimated by drawing a sample from the posterior distribution using the reparameterisation trick [20].

C Implementation Detail

C.1 DCDI and Graph Learning

This section covers the formulation of DCDI [8] and the graph learning method. These are subsequently used in the learning of VCD.

Given samples from an observed data distribution $P_X^{(0)}$ and K intervened distributions $P_X^{(k)}$, DCDI optimises a probabilistic belief over causal graphs $\mathcal{G}$ and intervention targets $\mathcal{I}$. Specifically, these are encoded as random binary matrices, $M_\mathcal{G}$ and $R_\mathcal{I}$, where $M_{ij}^\mathcal{G} = 1$ implies that the edge (i, j) is in the causal graph, and $R_{ki}^\mathcal{I} = 1$ implies that the variable x_i is in the intervention targets in environment k. Each entry in $M^\mathcal{G}$ follows an independent Bernoulli distribution, parameterised by matrix α where $P(M_{ij}^\mathcal{G} = 1) = \sigma(\alpha_{ij})$. $R_{ki}^\mathcal{I}$ is similarly parameterised by β. Under this parameterisation, causal discovery can be formulated as maximising the expected data log likelihood with sparsity regularisation,

$$L(\theta, \alpha, \beta) = \mathbb{E}_{\alpha,\beta}\left[\sum_{k=0}^{K} log[p_\theta^{(k)}(x_1^k, ..., x_d^k; \mathcal{G}, \mathcal{I})] - \lambda_G|\mathcal{G}| - \lambda_I|\mathcal{I}|\right], \tag{18}$$

where $p^{(k)}$ is the data likelihood under causal graph $\mathcal{G}$ and intervention targets $\mathcal{I}$ in the kth environment, as factorised in eq.(4). The conditional distributions are parameterised as feedforward neural networks with parameter θ; $\lambda_{G,I}$ are hyperparameters to control sparsity. In the original DCDI framework, this is also subject to an acyclicity constraint. However, this is not neccessary in the context of our work as we assume there are no instantaneous causal effects (i.e., within a timestep).

The training objective for VCD can be viewed as a modified version of the DCDI objective, where the likelihood term is replaced with the ELBO (Eq. 12),

$$L^{VCD}(\theta, \phi, \alpha, \beta) = \mathbb{E}_{\alpha,\beta}\left[\sum_{k=0}^{K}\sum_{t=0}^{T} \mathbb{ELBO}(\theta, \phi; \mathcal{G}, \mathcal{I}) - \lambda_G|\mathcal{G}| - \lambda_I|\mathcal{I}|\right], \tag{19}$$

Note that the expected number of edges in $\mathcal{G}$ and $\mathcal{I}$ given α and β is simply the sum of the probability of each entry being one. Therefore, the training objective can be computed as:

$$L^{VCD}(\theta, \phi, \alpha, \beta) = \mathbb{E}_{\alpha,\beta}\left[\sum_{k=0}^{K}\sum_{t=0}^{T} \mathbb{ELBO}(\theta, \phi; \mathcal{G}, \mathcal{I})\right] - \lambda_G \sum_{ij} \sigma(\alpha_{ij}) - \lambda_I \sum_{ki} \sigma(\beta_{ki}). \tag{20}$$

The gradients through the outer expectation can be estimated using the Gumbel-Softmax trick [18]. To implement this, the ELBO term is evaluated with a sample of the causal graph using the following expression for each entry,

$$M_{ij}^\mathcal{G} = \mathbb{I}(\sigma(\alpha_{ij} + L_{ij}) > 0.5) + \sigma(\alpha_{ij} + L_{ij}) - stop_gradient(\sigma(\alpha_{ij} + L_{ij})), \tag{21}$$

where $\mathbb{I}(\cdot)$ is the indicator function, L_{ij} is a sample from the logistic distribution, and $stop_gradient$ is a function that does not change the value of the argument but sets the gradient to zero. Samples for the intervention targets are similarly acquired. Note that the sample is used throughout each trajectory, i.e. the same sample graph and intervention targets are used for all of T timesteps.

C.2 Model Architecture

The encoders and decoders are parameterised as convolutional and deconvolutional networks from [12]. In the RSSM models, the transition models are parameterised as feedforward MLPs with two hidden layers of 300 hidden units. The recurrent module is a GRU with 300 hidden units. In VCD, to compensate for the fact each dimension in the latent space has a separate model, the number of hidden units in the GRU and MLP are reduced to 32 to avoid over-parameterisation. We found that initialising the encoders and decoders by pretraining them as a variational autoencoder helped with training stability for both RSSM and VCD.

The training objective is maximised using the ADAM optimiser [21] with learning rate 10^{-3} for mixed-state, and 10^{-4} for images. In both environments, we clip the log variance to -3, with a batch size of two trajectories from each of six environments with $T = 50$. In VCD, the hyperparameters $\lambda_\mathcal{G}, \lambda_\mathcal{I}$ are both set to 0.01. All models are trained on a single Nvidia Tesla V100 GPU.

D Qualitative Exploration for Learnt Latent Space

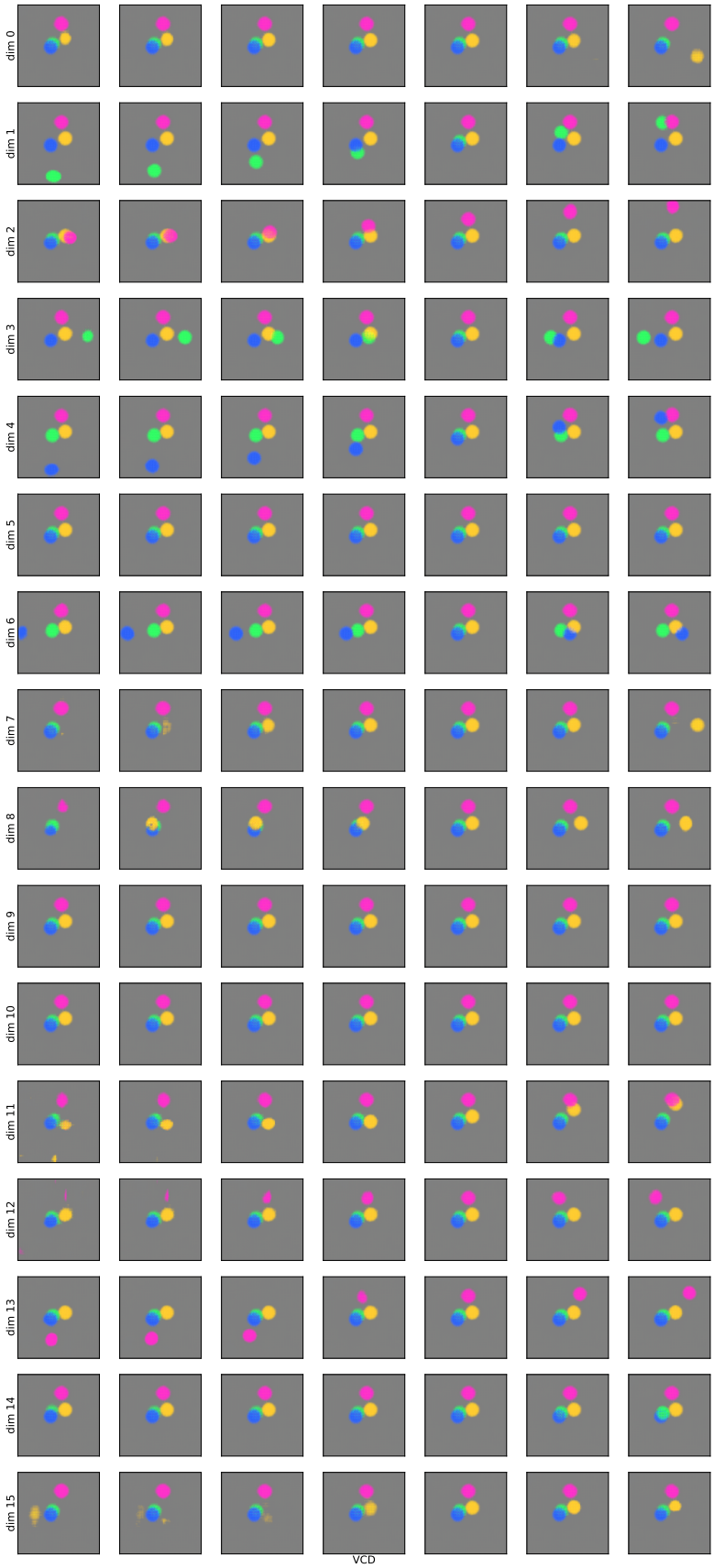


Figure 4: Visualisation of the learnt latent space in VCD.

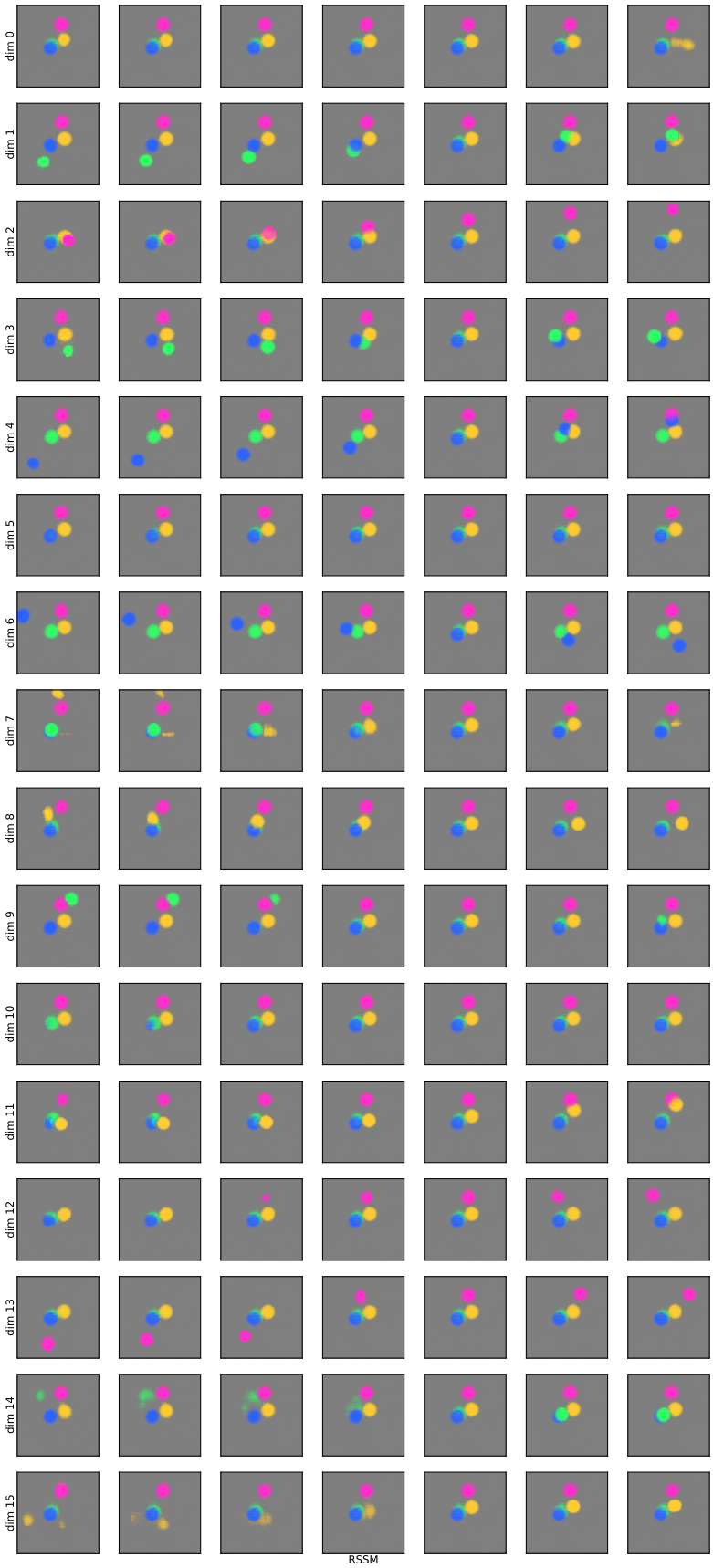


Figure 5: Visualisation of the learnt latent space in RSSM.