Learning Variational Latent Dynamics: Towards Model-based Imitation and Control

Hang Yin[1,2], Francisco S. Melo[1], Aude Billard[2] and Ana Paiva[1]

***Abstract*— In this paper, we learn dynamics from high-dimensional demonstrations to facilitate model-based prediction and robot control. The proposed approach leverages the progress in variational-bayes and sequence modeling, extracting a low-dimensional latent space so the dynamical relations of interest can be compactly represented and learned. Different from existing works, our model captures latent dynamics in a more general form and features efficient inference for pattern filtering, prediction and synthesis. The extracted feature mapping and latent dynamics can be naturally integrated in robot learning, yielding task imitation from raw data and prediction-based reproduction. The performance of latent dynamics learning and model-based imitation is shown in three tasks: 1) reconstructing and predicting images of bouncing balls movement with an accuracy competitive to the state-of-the-art; 2) synthesizing diverse handwriting image sequences; 3) learning to strike a ball under partial visual input, with results significantly outperforming baselines.**

I. INTRODUCTION

The capability of learning complex sensorimotor skills is one of the hallmarks of robots with a greater autonomy and a potential to work in unstructured environments. To achieve this, robots often need to process rich perceptions. One illustrative example can be found in Figure 1: a robot observes a ball rolling down a slope and moves its arm in an attempt to goal-strike. Predicting the dynamics of ball movement is necessary for the robot to prepare its motor command early enough to strike on time. One challenge is that the sensory input, in our case a raw video stream, is very high-dimensional and only a tiny fraction of this input is informative (the pixels that pertain to the ball position). Existing approaches to learn models of dynamical systems have relied on low-dimensional examples, where input explicitly represents the state of system dynamics [1], [2], [3]. However, in the case considered here, although the visual representation is of a high-dimensional unstructured form, the underlying process is largely governed by a low-dimensional model describing the motion of the ball. We thus investigate how to extract this low-dimensional space and establish the dynamical relations therein.

We build upon recent progresses on modeling high-dimensional sequential patterns [4], [5]. In particular, the framework based on variational-bayes and representation learning introduces latent variables, enabling a tractable optimization of the bound of original data likelihood. The latent variables are associated with observations through the posterior distribution model, whose design plays a key role in the variational framework. Existing works [6], [7] tend to perform the posterior inference upon future observations, which have merits in contexts such as language processing. They are, however, not applicable to robotics, where inference must be performed on-line as observations come.

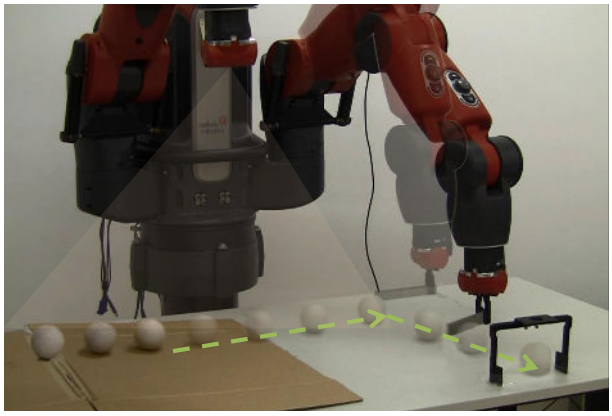


Fig. 1: Learning a motor task which requires reasoning about unstructured sensory inputs and their underlying dynamical relations: A robot strikes a ball based on the stream of images and the predicted states (transparent balls).

Moreover, and returning to our rolling ball example, the estimated model for the ball movement implies the possibility of a model-based control. This could be vital for handling more challenging conditions, such as the delay or malfunction of sensory input. For instance, humans can anticipate the movement of a ball and intercept it based on an internal prediction without observing the full trajectory. Yet, reasoning about unstructured data via representation learning is relatively less explored in model-based learning and control paradigms.

In this paper, we propose to encapsulate high-dimensional sequences with dynamical relations in a space with a lower dimensionality. Such a dynamical model is then utilized in model-based task imitation and reproduction. Our approach embeds high-dimensional sensations with a posterior model that is more suitable for robotics contexts comparing with [6], [7]. The posterior estimation is designed as part of the latent variables and the enforced dynamical relation is of a general nonlinear form, different from the specific design in [8], [9]. Conceptually, these proposed structures yield a state space model for the prior dynamics, which is not considered in works like [5]. At the same time, the training loss is derived as a valid bound of marginal data likeli-

[1]Hang Yin, Francisco S. Melo and Ana Paiva are with GAIPS, INESC-ID and Instituto Superior Técnico, Universidade de Lisboa, Portugal

[2]Hang Yin and Aude Billard are with Learning Algorithms and Systems Laboratory, École Polytechnique Fédérale de Lausanne, Switzerland

hood, enjoying advantages comparing with other variational embedding works like [10]. Furthermore, we propose and demonstrate that the obtained posterior and prior dynamics, as useful intermediate models, can be utilized to reduce task data dimension and to enable control upon foreseen sensations. Shortly, in this work, we are not only interested in learning to model sequential sensory patterns, but also take a step towards latent state filtering, predicting, and as such, *acting under sensory uncertainties*. The paper demonstrates the contributed approach in two sequential pattern synthesis tasks and the motivating ball-striking task, to highlight its performance on dynamics learning and model-based control in the robotics context. To summarize, the main paper contributions include:

- A type of variational model with a general latent dynamics that facilitates efficient learning and inference upon rich, high-dimensional and unstructured sequential data.
- A model-based paradigm that analyzes raw visual inputs and latent dynamics, enabling robust reproduction in simulated and real-world ball striking tasks.

II. RELATED WORK

The advancement of representation learning evokes researchers' recent interests on modeling high-dimensional sequential data without explicit or handcrafted features. In [5], a model based on a similar variational recurrent posterior is shown to be effective in a few pattern synthesis tasks. However, [5] chooses to involve raw features in the transition of prior model, resulting in a non-state-space model and difficulties in applying to model-based control. [6] and [7] regularize the training loss with priors that are independent of raw observations but the proposed posterior model is defined over the full data sequence. This implies a richer posterior estimation and is preferred for smoothing tasks. Our approach follows a similar variational treatment while with a posterior model over the history, which is necessary in robot motor tasks. In other similar works, [8] and [9] do not rely on future observations and the observation history is embedded as the parameters of a mixture of locally-linear dynamical systems. Interesting results emerge from this informative structure, demonstrating a correlation between identified latent variables and physical states. In our work, the posterior estimation is part of the latent variables, whose dynamical relation is formulated as general nonlinear dynamics. From this perspective, our work generalizes the enforced dynamics, which could be more compact for capturing diverse sequential patterns as is show in results IV-B.

Meanwhile, many robotic tasks face challenges of curse of dimensionality of the task data, for which dimension reduction techniques are often applied. Successful applications include [11] and [12], where linear projections and discrete latent variables are considered. They also assume global latent variables without a dynamical structure. This is similar to [13] which although adopts continuous latent variables and nonlinear embeddings. [14] uses variational auto-encoders to first obtain a latent representation and then fit recurrent networks to capture the dynamical behavior. Our approach learns the recognition and dynamics model in joint manner and more importantly, adopts a richer posterior family, handling the raw data without a Markovian assumption. Variational latent dynamics is successfully employed in [10] in a simulated inverted-pendulum example. Differently, as is pointed in [8], the dynamics training is not performed with a well-defined likelihood surrogate and requires i.i.d. data frames like [14]. Similar to [10], [15] realizes a planning-based control by reasoning about raw images with neural models. The models, however, are not learned as generative ones with a variational treatment.

III. APPROACH

A. Variational Latent Dynamics

We are interested in building a probabilistic model of high-dimensional sequences $\boldsymbol{x}_{0:T} \in \Re^{N_x \times T}$, with the dynamical relations captured in a compact latent space through the pair of latent variables $\boldsymbol{z}_{0:T} \in \Re^{N_z \times T}$ and $\boldsymbol{h}_{0:T} \in \Re^{N_h \times T}$, with $N_z = N_h \ll N_x$. Considering the log-likelihood of $\boldsymbol{x}_{0:T}$, a variational evidence lower bound (ELBO) can be derived:

$$\begin{aligned}\log p(\boldsymbol{x}_{0:T}) > \mathbb{E}_{q(\boldsymbol{z}_{0:T},\boldsymbol{h}_{0:T}|\boldsymbol{x}_{0:T})}[\log p(\boldsymbol{x}_{0:T}|\boldsymbol{z}_{0:T},\boldsymbol{h}_{0:T})] \\ - \mathrm{KL}[q(\boldsymbol{z}_{0:T},\boldsymbol{h}_{0:T}|\boldsymbol{x}_{0:T})\|p_0(\boldsymbol{z}_{0:T},\boldsymbol{h}_{0:T})]\end{aligned} \tag{1}$$

where $\boldsymbol{z}_{0:T}$ and $\boldsymbol{h}_{0:T}$ denote the sequence of latent variables, with q and p_0 as approximated posterior and prior distributions in Equation (1) respectively. We decompose the latent representation into two parts: 1) $\boldsymbol{z}_t$ as a probabilistic encoding to fit the variational inference framework; 2) $\boldsymbol{h}_t$ as a deterministic dynamics prediction (see below). Taking a first-order dynamical system perspective on the latent variables, we assume an observation only depends on its history and $\{\boldsymbol{z}_t, \boldsymbol{h}_t\}$ are a sufficient representation, namely:

$$\begin{aligned}&\log p(\boldsymbol{x}_{0:T}) - \log p(\boldsymbol{x}_0) = \sum_{t=1}^{T} \log p(\boldsymbol{x}_t|\boldsymbol{x}_{0:t-1}) \\ >&\sum_{t=1}^{T}\{\int \log p(\boldsymbol{x}_t|\boldsymbol{z}_t,\boldsymbol{h}_t) q(\boldsymbol{z}_t,\boldsymbol{h}_t|\boldsymbol{x}_{0:t}) d\boldsymbol{z}_t d\boldsymbol{h}_t \\ &-\mathrm{KL}[q(\boldsymbol{z}_t,\boldsymbol{h}_t|\boldsymbol{x}_{0:t})\|p_0(\boldsymbol{z}_t,\boldsymbol{h}_t|\boldsymbol{z}_{t-1},\boldsymbol{h}_{t-1})]\}\end{aligned} \tag{2}$$

Both posterior q and prior p_0 are further factorized into two parts, introducing an interaction between the latent variables $\boldsymbol{z}$ and $\boldsymbol{h}$:

$$\begin{aligned}q(\boldsymbol{z}_t,\boldsymbol{h}_t|\boldsymbol{x}_{0:t}) &= q(\boldsymbol{z}_t|\boldsymbol{h}_t,\boldsymbol{x}_t) q(\boldsymbol{h}_t|\boldsymbol{x}_{0:t-1}) \\ p_0(\boldsymbol{z}_t,\boldsymbol{h}_t|\boldsymbol{z}_{t-1},\boldsymbol{h}_{t-1}) &= p_0(\boldsymbol{z}_t|\boldsymbol{h}_t) p_0(\boldsymbol{h}_t|\boldsymbol{z}_{t-1},\boldsymbol{h}_{t-1})\end{aligned} \tag{3}$$

Thus the latent embeddings are retrieved from the observations till now, unlike [6], [7] which need to access future observations. To effectively back-propagate stochastic gradients, a deterministic temporal interaction is desired [5], [8]. Such an interaction can be shared between q and p_0 since we expect posterior and prior models to follow an identical dynamics. Concretely, when taking samples from $q(\boldsymbol{h}_t|\boldsymbol{x}_{0:t-1})$ and $p_0(\boldsymbol{h}_t|\boldsymbol{z}_{t-1},\boldsymbol{h}_{t-1})$, we assume $\boldsymbol{h}_t$ can be obtained through a deterministic latent dynamics, given $\boldsymbol{h}_{t-1}$ and $\boldsymbol{z}_{t-1}$:

$$\boldsymbol{h}_t = f(\boldsymbol{h}_{t-1}, \boldsymbol{z}_{t-1}) \tag{4}$$

where f denotes a general nonlinear function which can be estimated, for instance, by a recurrent neural network or other nonlinear estimators [16]. The stochasticity of the process is embedded in taking samples from a prior $p_0(\boldsymbol{z}_t|\boldsymbol{h}_t)$, which is expected to match the estimation from $q(\boldsymbol{z}_t|\boldsymbol{h}_t, \boldsymbol{x}_t)$ (KL term in Equation (1)) for a robust prediction of latent variables in face of a partial observable $\boldsymbol{x}_{0:T}$. Figure 2(a) and 2(b) illustrate inference for generation and recognition as graphical models.

B. Training and Inference

We parameterize the likelihood surrogate in a way similar to the popular variational-bayes method [17], [18], which approximates all of the conditional distributions of interest in Equation (2) as Gaussians. The conditioned variables are mapped to the means and diagonal covariance matrices through nonlinear function approximators. In light of the deterministic dynamics of Equation (4), a one-layer Long Short Term Memory (LSTM) is used to mitigate the vanishing gradients in back-propagation through the time [19]. The training objective can be optimized with respect to parameters for generative model $\boldsymbol{\theta}_g$, recognition model $\boldsymbol{\theta}_r$, prior $\boldsymbol{\theta}_0$ and latent dynamics $\boldsymbol{\theta}_f$:

$$\begin{aligned}&\mathcal{L}(\boldsymbol{\theta}_g, \boldsymbol{\theta}_r, \boldsymbol{\theta}_0, \boldsymbol{\theta}_f) = \sum_{t=1}^{T}\{ \\ &\mathbb{E}_{q_{\boldsymbol{\theta}_r}(\boldsymbol{z}_t|f_{\boldsymbol{\theta}_f}(\boldsymbol{h}_{t-1}, \boldsymbol{z}_{t-1}), \boldsymbol{x}_t)}[\log p_{\boldsymbol{\theta}_g}(\boldsymbol{x}_t|\boldsymbol{z}_t, f_{\boldsymbol{\theta}_f}(\boldsymbol{h}_{t-1}, \boldsymbol{z}_{t-1}))] \\ &-\mathrm{KL}[q_{\boldsymbol{\theta}_r}(\boldsymbol{z}_t|f_{\boldsymbol{\theta}_f}(\boldsymbol{h}_{t-1}, \boldsymbol{z}_{t-1}))\|p_{\boldsymbol{\theta}_0}(\boldsymbol{z}_t|f_{\boldsymbol{\theta}_f}(\boldsymbol{h}_{t-1}, \boldsymbol{z}_{t-1}))]\}\end{aligned} \tag{5}$$

The recognition model transforms sensory observations into a low-dimensional feature space. Note the encoding also depends on $\boldsymbol{h}_t$ which carries the information of previous observations to the current step. Synthesizing $\boldsymbol{x}_t$ given $\boldsymbol{x}_{0:t-1}$ can be achieved by recursively applying Equation (4), $p_0(\boldsymbol{z}_t|\boldsymbol{h}_t)$ and the generation model (Figure 2(a)):

$$p(\boldsymbol{x}_t|\boldsymbol{x}_{0:t-1}) = \int p(\boldsymbol{x}_t|\boldsymbol{z}_t, \boldsymbol{h}_t)p_0(\boldsymbol{z}_t|\boldsymbol{h}_t)q(\boldsymbol{h}_t|\boldsymbol{x}_{0:t-1})d\boldsymbol{z}_t\boldsymbol{h}_t \tag{6}$$

According to this equation, we can obtain an empirical estimation of $\boldsymbol{x}_t$ by conducting Monte-Carlo sampling.

C. Model-based Imitation Learning and Control

The models parameterized by $\boldsymbol{\theta}_r$ and $\boldsymbol{\theta}_f$ encode the original feature $\boldsymbol{x}_t$. We propose to leverage these models to learn tasks from expert demonstrations $\{\boldsymbol{x}_t, \boldsymbol{u}_t\}$. One way to realize this is to estimate a joint density as Gaussian Mixture Models (GMM). Here, in particular, the low-dimensional encodings $\boldsymbol{z}_t$ alleviate the curse-of-dimensionality and enable full covariance matrices. In reproduction, the control is derived by maximizing the conditional likelihood. To this end, the GMM can be viewed as an approximation of a Boltzmann distribution, resulting in a type of maximum entropy imitation learning [20]:

$$\begin{aligned}p(\boldsymbol{z}_t, \boldsymbol{h}_t, \boldsymbol{u}_t) &\approx \sum_{k=1}^{N} w_k \mathcal{N}(\boldsymbol{z}_t, \boldsymbol{h}_t, \boldsymbol{u}_t|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \\ &\propto p(\boldsymbol{u}_t|\boldsymbol{z}_t, \boldsymbol{h}_t) = \frac{e^{-Q(\boldsymbol{z}_t, \boldsymbol{h}_t, \boldsymbol{u}_t)}}{\int e^{-Q(\boldsymbol{z}_t, \boldsymbol{h}_t, \boldsymbol{u}_t)}d\boldsymbol{u}_t} \\ \boldsymbol{u}_t^* &= \underset{\boldsymbol{u}_t}{\operatorname{argmin}}\, Q(\boldsymbol{z}_t, \boldsymbol{h}_t, \boldsymbol{u}_t) = \underset{\boldsymbol{u}_t}{\operatorname{argmax}}\, p(\boldsymbol{u}_t|\boldsymbol{z}_t, \boldsymbol{h}_t)\end{aligned} \tag{7}$$

Here $Q(\cdot)$ denotes statistical moments that cause the expert demonstrations to incur a low cost and $\{w_k, \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k\}$ are the GMM parameters.

The latent dynamical system $p_0(\boldsymbol{z}_t, \boldsymbol{h}_t|\boldsymbol{z}_{t-1}, \boldsymbol{h}_{t-1})$ provides a prediction of latent variables to apply the learned controller free of sensory input. This adds an extra value for a potentially more robust and flexible task reproduction with a delayed or missed $\boldsymbol{x}_t$. The entire learning pipeline is schematically shown as Figure 3.

IV. IMPLEMENTATION AND RESULTS

A. Bouncing Balls

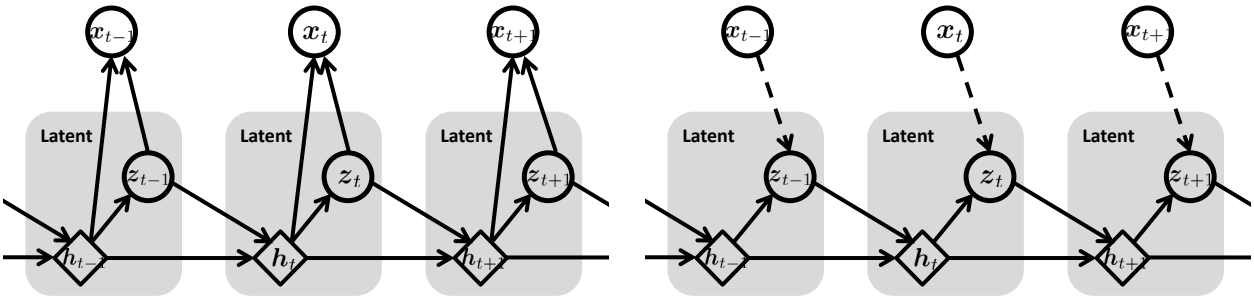
We first validate the approach to model the dynamics of a group of bouncing balls. We use the bouncing-ball dataset, a dataset of images showing 3 bouncing balls and used previously to assess dynamical patterns generators/predictors [21], [22]. We follow the supplementary script of [21] to generate 30×30 grayscale images, constructing 4000 sequences as the training data and 1000 as the test data. The samples of synthesized movement are demonstrated in the supplementary video.

Table I reports one-step pixel prediction error, with some of the results from [22]. Although our approach does not directly optimize prediction accuracy, it achieves an average error on par with previous approaches while using latent dynamics of a smaller size (the second column in Table I). One possible explanation to this is that the enforced nonlinear dynamics allow is flexible to capture the sequential data in a more compact form.

TABLE I: Pixel Prediction Error

Model	Latent Size	Pred. Err.
DTSBN-s	100-100	2.79 ± 0.39
TSBN-ORDER-4	100	3.07 ± 0.40
TSBN-ORDER-1	100	9.48 ± 0.38
RTRBM	3750	3.88 ± 0.33
SRTRBM	3750	3.31 ± 0.33
KVAE	16×9	3.79 ± 1.04
Ours	64	3.07 ± 0.89

In Figure 4, we present reconstructed and predicted images, where the reconstruction (the second row) from the filtered latent state turns out to match the true motion (the first row) well. The last two rows highlight reconstructed images from predicted latent variables. As was observed in [8], the prediction without sensory input (the third row, right of dash line) can be reliable only for a short time horizon. Eventually the predicted dynamics moves away from the true



(a) Graph of generative model (b) Graph of recognition model

Fig. 2: Graphical representation of the proposed model: The latent state is an augmentation of a stochastic component $\boldsymbol{z}$ (circular node) and $\boldsymbol{h}$ (diamond node) that has a deterministic dependency on the previous state. The latent state has an internal dependency between $\boldsymbol{z}$ and $\boldsymbol{h}$. The belief of $\boldsymbol{z}$ can be propagated through the latent dynamics, while it could also depend on $\boldsymbol{x}$ in the recognition model (Figure 2(b)). The dashed arrows mean the dependency on observation is optional so the estimation can be fully governed by the latent state dynamics.

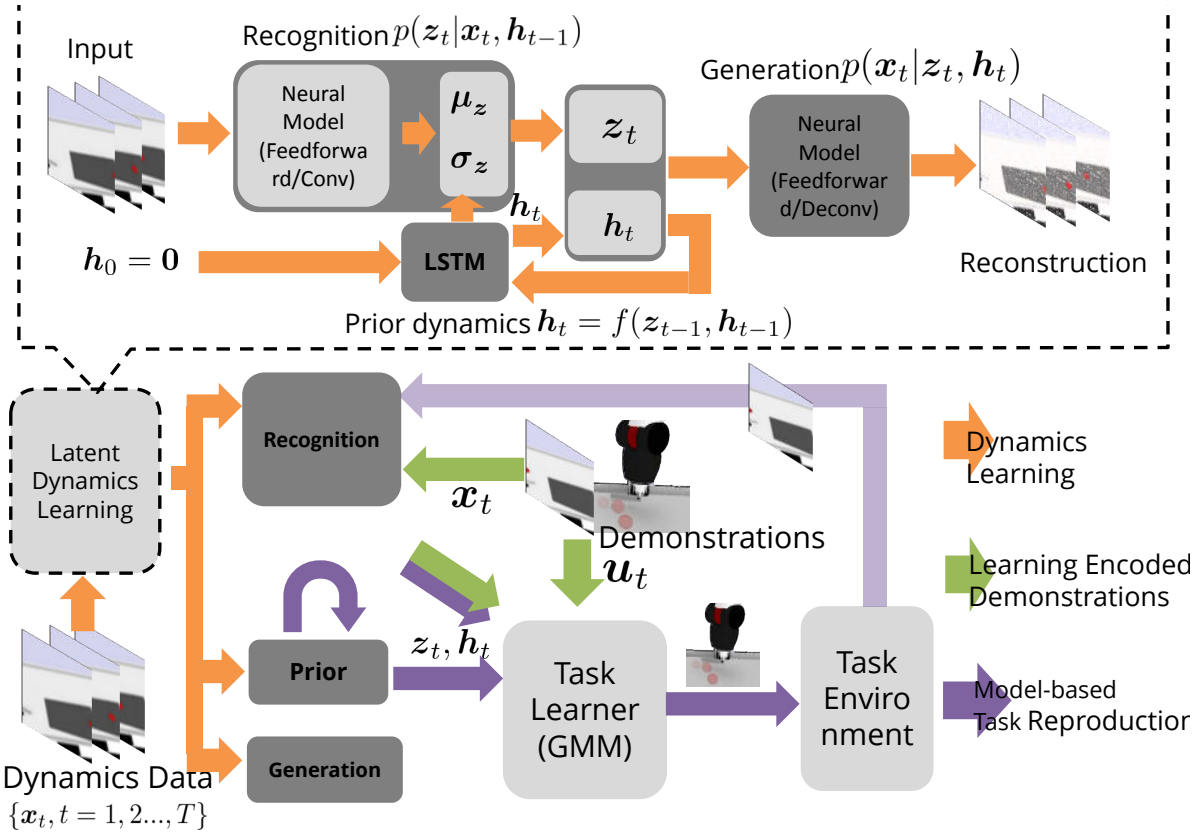


Fig. 3: A schematic view of pipelines: Models for generation, prior dynamics and feature recognition (shapes with a darker gray color) are obtained through latent dynamics learning in Section III-B (subprocess in the dashed frame); Demonstration data are encoded by the recognition model to build the task learner in the latent space (green flows); Reproduction (purple flows) reasons about raw observations with the recognition model or resorts to an internal prediction when observations are not available (transparent purple arrow).

trajectory. Long-term predictions retains only a qualitative accuracy, as is shown in Figure 4. A quantitative comparison can be found in Figure 5, where our approach consistently outperforms the baseline on pixel square errors.

B. Modeling Handwriting Image Sequences

In this experiment, the proposed approach is applied to modeling rich handwriting samples, which include alphabetical letters (capital and lower case) and digits from the UJI Char Pen 2 dataset [23]. The algorithm in [24] is used to diversify letter trajectories. These trajectories are then used

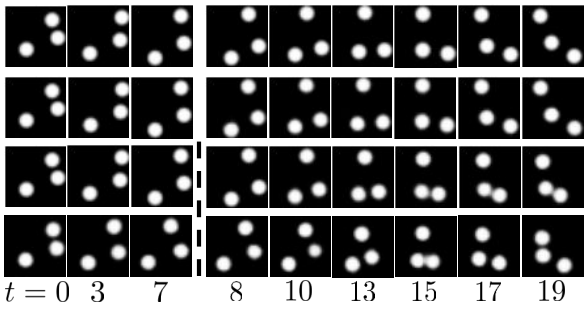


Fig. 4: Each row of samples: 1) ground truth; 2) reconstruction with input image (our approach); 3) and 4) reconstruction with the first 8 frames (left of dash line, t = 0, 3, 7) and prediction of remained 12 steps (right of dash line, t = 8, 10, 13, 15, 17, 19) Row 3) - ours. Row 4) - KVAE.

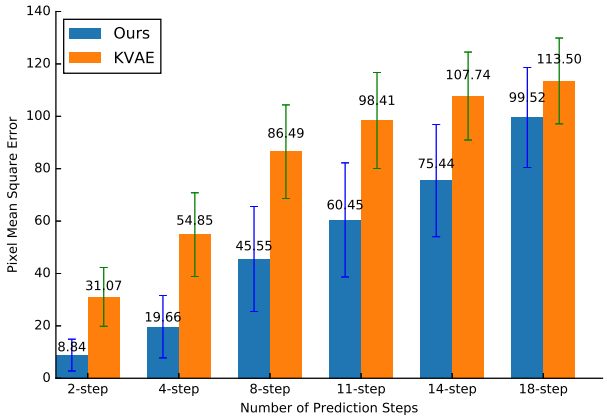


Fig. 5: Accuracy (mean square of pixel error across time horizon) of forward prediction as the number of time steps increase.

to generate 1984 video clips for 62 character types. Each sequence consists of 20 grayscale images of a 28×28 size.

The synthesis result is shown in Figure 6(a). The model successfully learns to generate plausible samples of different types with the same model setup as the bouncing ball example. Note that sometimes a fraction of the pixels are not exactly kept in the next frames. For instance, the initial step of generating "e" (the second column in Figure 6(a)) does not match the finalized image in terms of pixels. This is because the pixels are not incrementally filled and the latent dynamics only propagates abstract states for reconstructing the entire image. Still, the overall stroke direction and pattern are consistent throughout the generation process.

We find the KVAE approach at modeling locally-linear latent dynamics appears inadequate in this task, at least for models with a similar size. Most failing samples exhibit an oscillating pattern or an inconsistent generation process, as are shown in the lower part of Figure 6(a). This is probably due to the fact that, comparing with the bouncing-ball example, the dynamical behaviors here are much more diverse and the generation of pixels requires to encode a long-history dependency. Mixing locally-linear dynamics might not be sufficient to handle this without resorting to a vast number of local models.

Figure 6(b) presents some samples of completing the character formation with seeding images. Specifically, eight initial frames of the test samples are fed and the prior dynamics is recursively applied to generate the remaining steps. One can find that the synthesized sequences conform to the initial steps and complete the formation with reasonable characters.

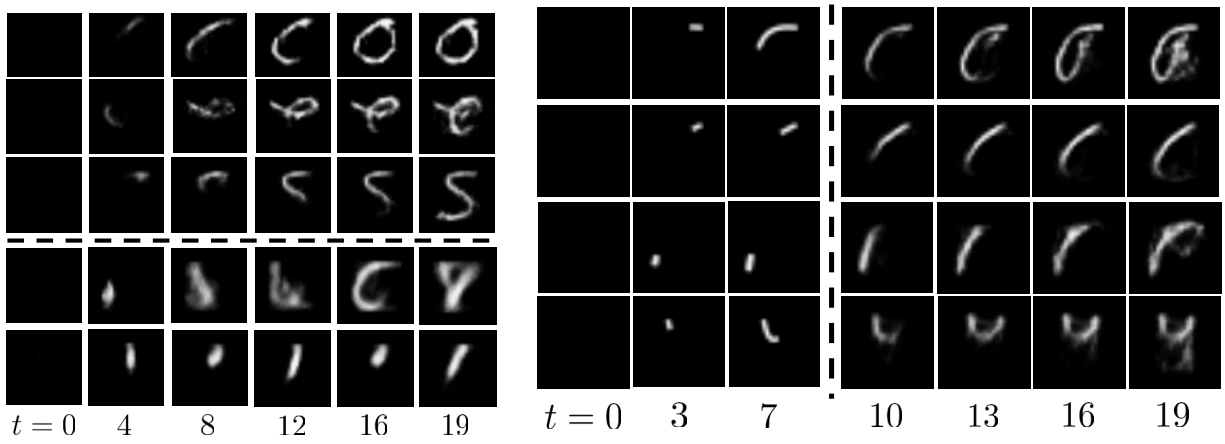
C. Learning to Strike a Rolling Ball

Finally, the proposed approach is applied to the motivating task illustrated in Figure 1. We teach a Baxter robot through tele-operation to learn from demonstrations to strike a rolling ball towards a goal. The task setup is shown as a Roboschool simulation environment [25] in Figure 7(a)[1]. The ball starts rolling on the slope with a random velocity. The robot moves a paddle to score a goal by striking the ball at a proper position and instant. The state of ball movement can only be observed through a camera and the real ball position and velocity are unknown to the robot. Meanwhile, the visual input is switched off after the first few frames so the robot has to anticipate and act based on the model prediction. To demonstrate the task, a human operator watches the ball rolling movement and uses a keyboard to steer the end-effector position and initiate the strike. The successful demonstrations record the paddle position (u_p) for goal-making strikes as well as the frames from the ball starts rolling until the robot strikes it ($\boldsymbol{x}_{0:t_s}$).

To learn the rolling dynamics, 50 rollouts with random initial velocities are collected without striking attempts. Each rollout lasts 150 simulation steps. The visual data are RGB images composed of $50 \times 50 \times 3$ pixel values. In addition, we demonstrate successful strikes under 16 perturbation conditions. The GMMs $p(\boldsymbol{x}_{t_s})$ and $p(\boldsymbol{x}_t, u_p)$ are learned to model, respectively, the density of striking frame and the joint density of images and robot action. In the test reproduction, the robot moves according to $p(u_p|\boldsymbol{x}_t)$ and exercises a strike through a torque controller when $p(\boldsymbol{x}_{t_s}) > \epsilon$. Here, ϵ is a threshold to decide if the current frame is sufficiently similar to demonstrations and its value is chosen in an ad-hoc manner as 10^{-3}. The task learning is performed under three settings:

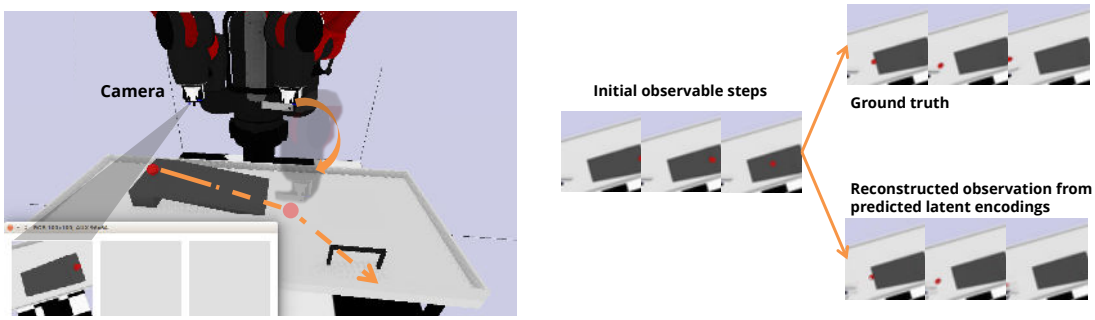
- **Full observability**: GMM models are learned from the 16 demonstrations with striking attempts, using real ball position and velocity. The real ball position and velocity are also always observable in the test runs. Hence no model prediction is performed in this condition.
- **Proposed approach with partial observability**: We first learn the recognition model $p(\boldsymbol{z}_t|\boldsymbol{x}_{0:t})$ and the prior latent dynamics $p_0(\boldsymbol{z}_t, \boldsymbol{h}_t|\boldsymbol{z}_{t-1}, \boldsymbol{h}_{t-1})$ from 50 rollouts without striking attempts. The dimension of combined latent variable is 8. GMM models are trained on the

[1] https://github.com/navigator8972/roboschool_baxterstriker



(a) (b)

Fig. 6: (a) Samples of synthesizing handwriting image sequences. The sampling starts from a zero latent variable and recursively applies the learned latent dynamics and a Gaussian random perturbation. Each row: steps (t = 0, 4, 8, 12, 16, 19) of the synthesized character sequence: upper - ours; lower - KVAE ([9]) (b) Samples of completing image sequences given the first few handwriting frames. The synthesis starts with a latent variable encoded from the pivoting frames. Each row: pivoting frames (left of the dash line, t = 0, 3, 7); completed image steps (right of the dash line, t = 10, 13, 16, 19).



(a) (b)

Fig. 7: (a) Task setup of Baxter ball striking domain: The robot observes the ball movement through a wrist camera and moves a paddler to strike the ball on the goal. The visual sensing can be off halfway so an internal model is desired to determine the possible trajectory (dash line) and its action. (b) Encoding and predicting what the robot observes for a strike demonstration: ground truth and reconstruction from the propagation of latent dynamics. The dynamics is learned from data without strike actions.

16 demonstrations whose images are encoded with the recognition model. In the test runs, after the ball starts rolling, the robot has access to images during the first 30/40/50/60 steps and uses the prior dynamics model to predict and act afterwards.

- **Using original feature or a linear latent representation**: Here we learn both dynamics and control policies without extracting the latent representation or using one obtained from linear dimension reduction. For dynamics learning, an additional GMM model is trained to model the adjacent observations pairs $p(\boldsymbol{x}_{t+1}, \boldsymbol{x}_t)$ in the 50 rollouts without striking attempts. The baseline of a linear latent representation is retrieved through a PCA explaining 99% data variance, reducing the original dimension from 7500 to 137.

Figure 7(b) shows that the predicted frames from the learned dynamical model match the real ones well. The main differences are about the paddle and ball position after the strike occurs. This is as expected because the behaviors of paddle and ball-striking are unseen to the model.

Figure 8 shows a quantitative assessment of performance in simulation and qualitative demonstration through a series of snapshots. For each learned controller, the robot starts with a fixed posture and is tested with 20 consecutive trials. The test is replicated five times so the robot can score 100 goals in maximum. Since the success of scoring the goal is sensitive to the striking position and time, we also use the occurrences of hitting the ball as a part of the evaluation. As a result, the robot under the full observability condition manages to hit the ball every time and strikes on the target in most of the trials. This is not surprising because the robot enjoys a perfect access to the true state and its performance

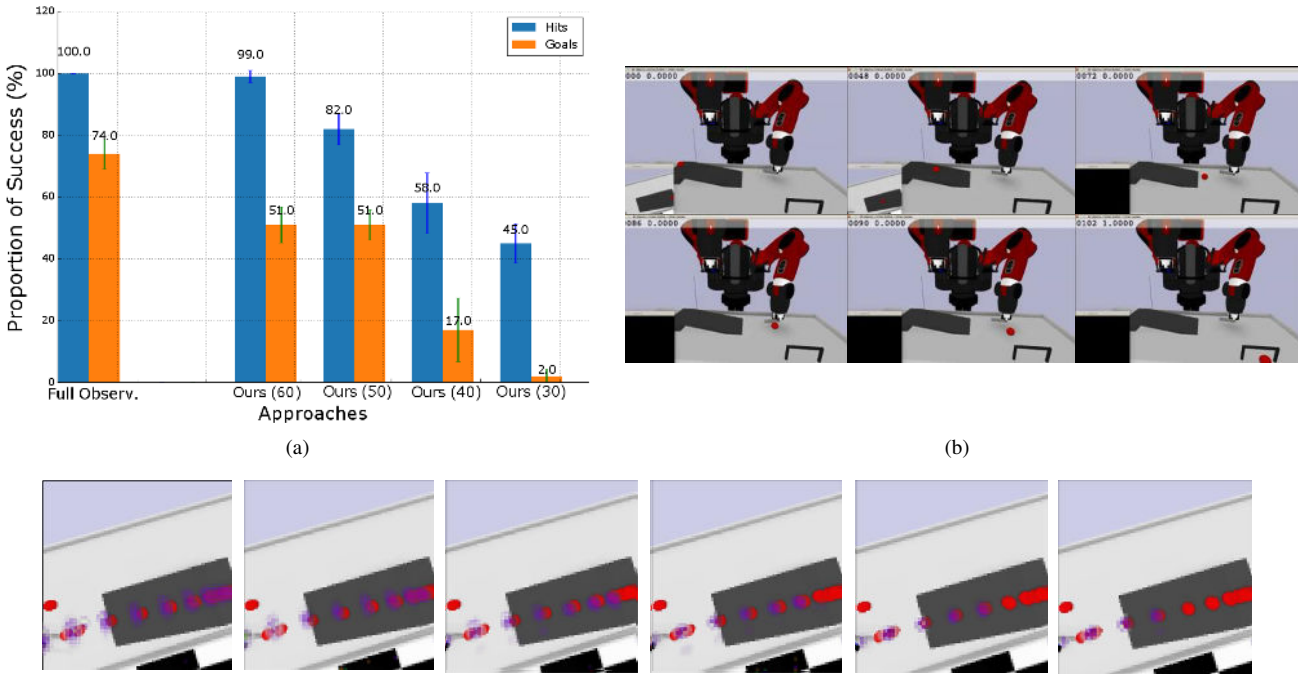


Fig. 8: Results of the Baxter ball striking example: (a) Proportion of successful hits and goals of different approaches and setups: The full observability setup can always retrieve the real state and no model prediction is performed. The number in the parentheses indicates the length of initial frames that are accessible to our approach in a partial observable setting. The baselines based on no/linear dimension reduction never hit the ball thus are not shown in the figure. (b) Snapshots of successful strikes from simulated Baxter control based on model prediction. (c) Prediction uncertainty by overlapping real and predicted ball position images with the time axis collapsed (t = 0, 9, 19, 29, 39, 49, 59, 69, 79, 81, 89): real - solid red color. The prediction - transparent purple color. The predicted motions are stochastic and 5 samples are taken and merged. From left to right: performing prediction 10/20/30/40/50/60 frames. The outlier of red dots indicates the position after striking so it is not close to predictions.

can be regarded as an upper-bound for the given imitation learning pipeline. Comparatively, the proposed variational dynamics learning can also secure quite a number of hits and goals, especially when the robot does not have to plan too long in the absence of sensory feedback. When the time window for observation is limited, e.g., only the first 30 frames are accessible, the latent dynamics could only provide a coarse prediction as the observations at early stage are ambiguous to determine an accurate future trajectory. This can be further demonstrated in Figure 8(c): the uncertainty of ball pixels gradually decreases as more frames can be observed before the model-based prediction. Therefore a rough prediction for a long horizon might help the robot to still hit the ball many times but is not accurate enough to ensure an on-goal strike. The results of no/linear latent representation baseline are omitted because of their poor performance: neither enables the robot to hit the ball once. This highlights the necessity of learning complex latent representations in handling unstructured sensory data, for which traditional approaches are inadequate or not directly applicable. Figure 8(b) and 9 demonstrate successful trials and the implementation on the simulated and real Baxter robot.

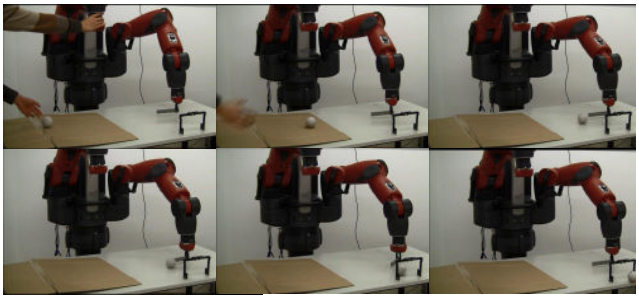


Fig. 9: Snapshots of successful strikes from Baxter control based on model prediction. Real camera data is collected and used for the real-world implementation.

V. CONCLUSION

The proposed approach is shown to be effective to extract a low-dimensional space and a dynamical relation, which enable efficient filtering/prediction of high-dimensional sequential patterns. Comparing with baseline methods, the adopted nonlinear embeddings and latent dynamics are advantageous in modeling diverse dynamical images and realizing an improved prediction accuracy. This facilitates learning and reproducing challenging visual motor tasks under a model-based setting, although care must be taken in determining the number of steps to predict ahead. Also, a large volume of data is desired to capture complex dynamical process, such as bouncing balls and handwriting formation, because the enforced dynamics is designed with limited task-relevant structure.

Our work opens several directions for further research. The latent dynamics learning is independent of the target task and the encoded dynamics exhibits certain robustness in dealing with untrained paddler behaviors. This motivates research towards a kind of task-agnostic learning, in which the robot first extracts useful features/constraints and then adapts to the target domain with limited task data. In terms of enforced dynamics, this paper adopts LSTM for an improved modeling flexibility. Actually, a more restrictive form which is convenient for analysis could also be considered for certain scenarios. For instance, one can try to impose a convergence guarantee for a goal-directed linear parameter varying system, hence extending works [26], [27]. Moreover, when demonstrating the task is not straightforward, the latent dynamics can be leveraged in the optimization of trajectory and policy [10], [15]. It is necessary to investigate how to utilize the unreliable long-term prediction in these settings, probably with additional information such as the uncertainty about latent variables.

ACKNOWLEDGMENT

This work is partially funded by Swiss National Center of Robotics Research and national funds through Fundação para a Ciência e a Tecnologia (FCT) with reference UID/CEC/50021/2013, the doctoral grant (SFRH/BD/51933/2012) under IST-EPFL Joint Doctoral Initiative and the project AMIGOS (PTDC/EEISII/7174/2014).

REFERENCES

[1] A. Paraschos, C. Daniel, J. Peters, and G. Neumann, "Probabilistic movement primitives," in *Proceedings of Neural Information Processing Systems (NIPS)*, 2013, pp. 2616–2624.

[2] S. Calinon, "Robot learning with task-parameterized generative models," in *Proceedings of the International Symposium of Robotics Research (ISRR)*, 2015.

[3] K. Kronander, S. M. Khansari Zadeh, and A. Billard, "Incremental motion learning with locally modulated dynamical systems," *Robotics and Autonomous Systems*, 2015.

[4] A. Graves, "Generating sequences with recurrent neural networks," *CoRR*, vol. abs/1308.0850, 2014.

[5] J. Chung, K. Kastner, L. Dinh, K. Goel, A. C. Courville, and Y. Bengio, "A recurrent latent variable model for sequential data," in *Proceedings of Neural Information Processing Systems (NIPS)*, 2015.

[6] M. J. Johnson, D. Duvenaud, A. B. Wiltschko, S. R. Datta, and R. P. Adams, "Composing graphical models with neural networks for structured representations and fast inference," in *Proceedings of Neural Information Processing Systems (NIPS)*, 2016.

[7] R. G. Krishnan, U. Shalit, and D. Sontag, "Structured inference networks for nonlinear state space models," in *Proceedings of the National Conference on Artificial Intelligence (AAAI)*, 2017, pp. 2101–2109.

[8] M. Karl, M. Soelch, J. Bayer, and P. van der Smagt, "Deep variational bayes filters: Unsupervised learning of state space models from raw data," in *Proceedings of the International Conference on Learning Representations (ICLR)*, 2017.

[9] M. Fraccaro, S. Kamronn, U. Paquet, and O. Winther, "A disentangled recognition and nonlinear dynamics model for unsupervised learning," in *Proceedings of Neural Information Processing Systems (NIPS)*. Curran Associates, Inc., 2017, pp. 3601–3610.

[10] M. Watter, J. T. Springenberg, J. Boedecker, and M. A. Riedmiller, "Embed to control: A locally linear latent dynamics model for control from raw images." *CoRR*, vol. abs/1506.07365, 2015.

[11] A. Colom, G. Neumann, J. Peters, and C. Torras, "Dimensionality reduction for probabilistic movement primitives," in *Proceedings of IEEE International Conference on Humanoid Robots (Humanoids)*, Nov 2014, pp. 794–800.

[12] E. Rueckert, J. Mundo, A. Paraschos, J. Peters, and G. Neumann, "Extracting low-dimensional control variables for movement primitives," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, 2015, pp. 1511–1518.

[13] H. Yin, F. S. Melo, A. Billard, and A. Paiva, "Associate latent encodings in learning from demonstrations," in *Proceedings of the National Conference on Artificial Intelligence (AAAI)*, San Francisco, USA, 2017.

[14] K. Sasaki, K. Noda, and T. Ogata, "Visual motor integration of robot's drawing behavior using recurrent neural network," *Robotics and Autonomous Systems*, vol. 86, pp. 184–195, 12 2016.

[15] C. Finn and S. Levine, "Deep visual foresight for planning robot motion," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, 2017, pp. 2786–2793.

[16] S. Kim and A. Billard, "Estimating the non-linear dynamics of free-flying objects," *Robotics and Autonomous Systems*, vol. 60, p. 11081122, 09 2012.

[17] D. P. Kingma and M. Welling, "Stochastic gradient vb and the variational auto-encoder," in *Proceedings of the International Conference on Learning Representations (ICLR)*, 2014.

[18] D. J. Rezende, S. Mohamed, and D. Wierstra, "Stochastic backpropagation and approximate inference in deep generative models," in *Proceedings of the International Conference on Machine Learning (ICML)*, 2014.

[19] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computing*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997.

[20] B. D. Ziebart, A. L. Maas, J. A. Bagnell, and A. K. Dey, "Maximum entropy inverse reinforcement learning." in *Proceedings of the National Conference on Artificial Intelligence (AAAI)*, 2008, pp. 1433–1438.

[21] I. Sutskever, G. E. Hinton, and G. W. Taylor, "The recurrent temporal restricted boltzmann machine," in *Proceedings of Neural Information Processing Systems (NIPS)*, 2009, pp. 1601–1608.

[22] R. H. D. C. Z. Gan, C. Li and L. Carin, "Deep temporal sigmoid belief networks for sequence modeling," in *Proceedings of Neural Information Processing Systems (NIPS)*, 2015.

[23] D. Llorens, F. Prat, A. Marzal, J. Vilar, M. Castro, J. Amengual, S. Barrachina, A. Castellanos, S. Espaa, J. Gmez, J. Gorbe, A. Gordo, V. Palazn, G. Peris, R. Ramos-Garijo, and F. Zamora, "The ujipenchars database: a pen-based database of isolated handwritten characters," in *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)*, Marrakech, Morocco, may 2008.

[24] H. Yin, P. Alves-Olivera, F. S. Melo, A. Billard, and A. Paiva, "Synthesizing robotic handwriting motion by learning from human demonstrations," in *Proceedings of International Joint Conference on Artificial Intelligence (IJCAI)*, New York, USA, 2016.

[25] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal policy optimization algorithms," *CoRR*, vol. abs/1707.06347, 2017.

[26] S. M. Khansari-Zadeh and A. Billard, "Learning Stable Non-Linear

Dynamical Systems with Gaussian Mixture Models," *Transactions on Robotics*, 2011.

[27] J. R. Medina and A. Billard, "Learning stable task sequences from demonstration with linear parameter varying systems and hidden markov models," in *Proceedings of Machine Learning Research: Conference on Robot Learning (CoRL)*, 2017, pp. 175–184.