ODE^{2}VAE: Deep generative second order ODEs with Bayesian neural networks

Aalto University School of Science

Çağatay Yıldız[1], Markus Heinonen[1,2], Harri Lähdesmäki[1]   [1]Aalto University, Finland   [2]Helsinki Institute for Information Technology

TL;DR

ODE^{2}VAE simultaneously learns the embedding of high dimensional trajectories via variational auto-encoders, and infers arbitrarily complex continuous-time latent dynamics. Our Bayesian neural network differential function prevents overfitting in low-data regimes and 2nd ODE system leads to more accurate predictions compared to 1st order ODEs.

Previous Work and Contributions

Existing techniques:

- VAEs are mostly for **static** data like images [1].
- RNN-based VAEs are discrete and fail to produce accurate **long-term** forecasts.
- Neural ODEs are **first-order**, with no regularization [2].

$$\dot{\mathbf{z}}_t := \frac{d\mathbf{z}_t}{dt} = \underbrace{\mathbf{f}(\mathbf{z}_t)}_{\text{NN}}, \qquad \mathbf{z}_T = \mathbf{z}_0 + \int_0^T \mathbf{f}(\mathbf{z}_t)dt$$

We propose:

- Second-order ODE with **position $\mathbf{s}_t$** and **momentum $\mathbf{v}_t$** latent spaces.
- **Probabilistic ODEs** with Bayesian neural network differential function.

$$\begin{aligned} \dot{\mathbf{s}}_t &= \mathbf{v}_t \\ \dot{\mathbf{v}}_t &= \underbrace{\mathbf{f}_{\mathcal{W}}(\mathbf{s}_t,\mathbf{v}_t)}_{\text{Bayesian NN}} \end{aligned} \qquad \begin{bmatrix}\mathbf{s}_T\\ \mathbf{v}_T\end{bmatrix} = \begin{bmatrix}\mathbf{s}_0\\ \mathbf{v}_0\end{bmatrix} + \int_0^T \begin{bmatrix}\mathbf{v}_t\\ \mathbf{f}_{\mathcal{W}}(\mathbf{s}_t,\mathbf{v}_t)\end{bmatrix} dt$$

Variational Inference

Denote by $\mathbf{x}_{0:N}$ an observed sequence, $\mathbf{z}_t := (\mathbf{s}_t, \mathbf{v}_t)$ combined state, $\mathcal{W}$ weights of differential function $\mathbf{f}_{\mathcal{W}}(\mathbf{s},\mathbf{v})$. Since the exact posterior is intractable, we resort to following variational posterior:

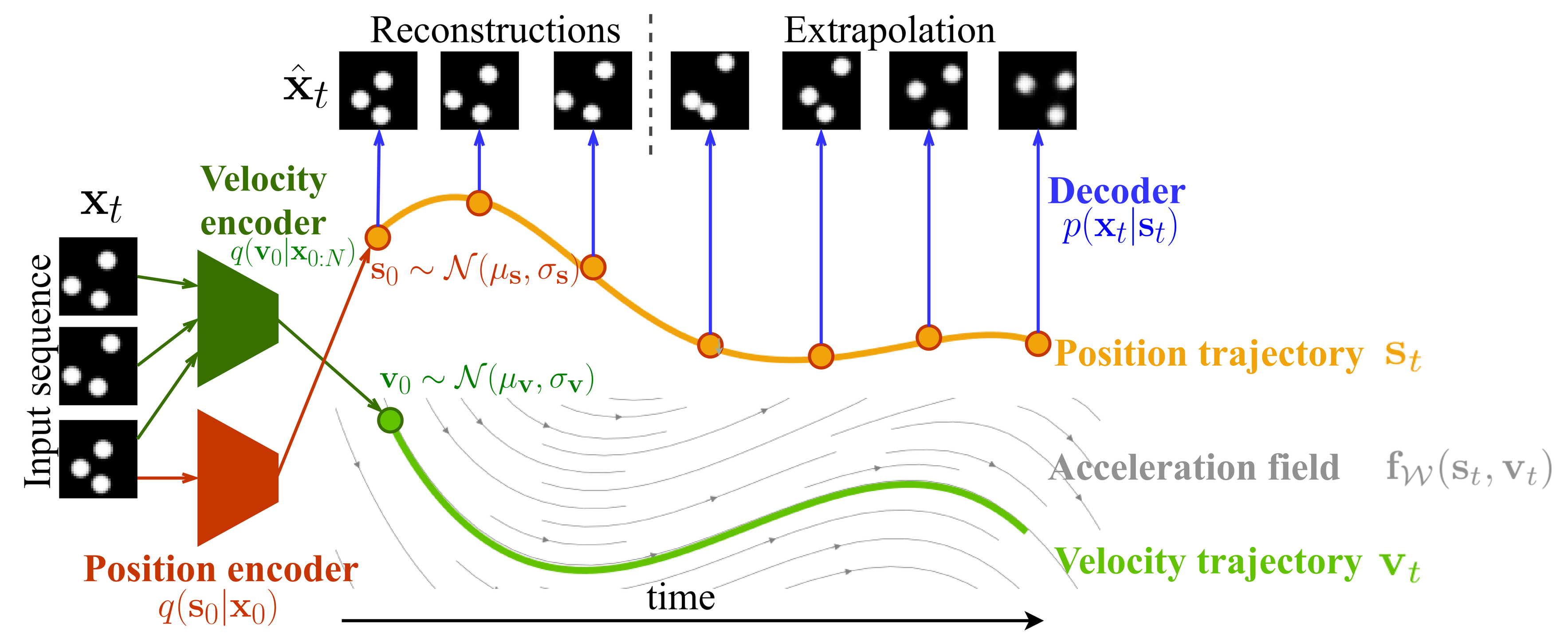
$$q(\mathcal{W}, \mathbf{z}_{0:N}|\mathbf{x}_{0:N}) = q(\mathcal{W})\; q_{\text{enc}}(\mathbf{z}_0|\mathbf{x}_{0:N})\; q_{\text{ode}}(\mathbf{z}_{1:N}|\mathbf{x}_{0:N}, \mathbf{z}_0, \mathcal{W})$$

The first two terms on rhs are Gaussian distributions. The last term follows another integral and can be computed by the instantaneous change of variables formula [2]. Then, ELBO becomes

$$\begin{aligned} \mathcal{L} = &\, \mathrm{KL}[q(\mathcal{W})||p(\mathcal{W})] && \texttt{\# ODE penalty} \\ &- \mathrm{KL}[q_{\text{enc}}(\mathbf{z}_0|\mathbf{x}_{0:N})||p(\mathbf{z}_0)] && \texttt{\# VAE penalty} \\ &- \sum_i \mathbb{E}_{q(\mathcal{W})}[\mathrm{KL}[q_{\text{ode}}(\mathbf{z}_i|\mathcal{W},\mathbf{x}_{0:N})||p(\mathbf{z}_i)]] && \texttt{\# dynamic penalty} \\ &+ \mathbb{E}_{q_{\text{enc}}(\mathbf{z}_0|\mathbf{x}_{0:N})}[\log p(\mathbf{x}_0|\mathbf{z}_0)] && \texttt{\# VAE reconstr.} \\ &+ \sum_i \mathbb{E}_{q(\mathcal{W})}\mathbb{E}_{q_{\text{ode}}(\mathbf{z}_i|\mathbf{x}_{0:N},\mathbf{z}_0,\mathcal{W})}[\log p(\mathbf{x}_i|\mathbf{z}_i)] && \texttt{\# dynamic reconstr.} \end{aligned}$$

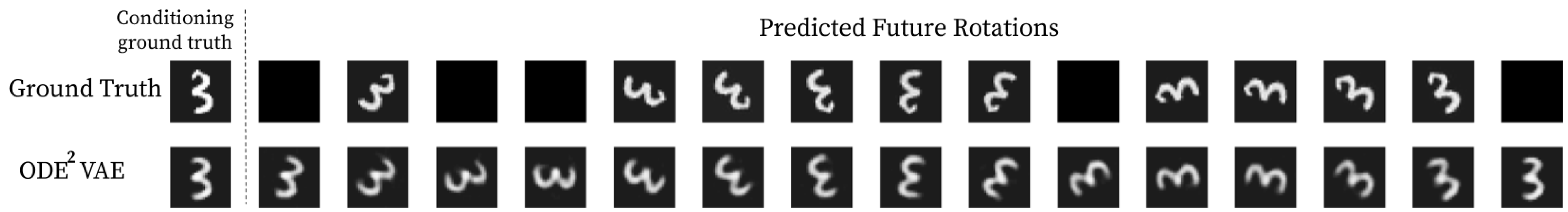
Generative Model

$$\begin{aligned} \mathbf{s}_0 &\sim p(\mathbf{s}_0) && \texttt{\# position initial value} \\ \mathbf{v}_0 &\sim p(\mathbf{v}_0) && \texttt{\# velocity initial value} \\ \mathbf{s}_t &= \mathbf{s}_0 + \int_0^t \mathbf{v}_\tau d\tau && \texttt{\# position trajectory} \\ \mathbf{v}_t &= \mathbf{v}_0 + \int_0^t \mathbf{f}_{\text{true}}(\mathbf{s}_\tau, \mathbf{v}_\tau) d\tau && \texttt{\# velocity trajectory} \\ \mathbf{x}_i &\sim p(\mathbf{x}_i|\mathbf{s}_i), \quad i \in [0, N] && \texttt{\# decoder} \end{aligned}$$



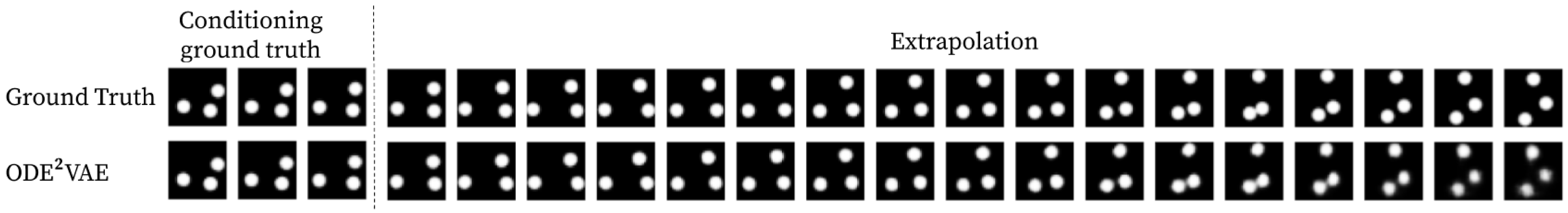
Rotating MNIST Experiment

Dataset: Sequences of rotating handwritten 3s with 30% missing data (black squares below) [?].



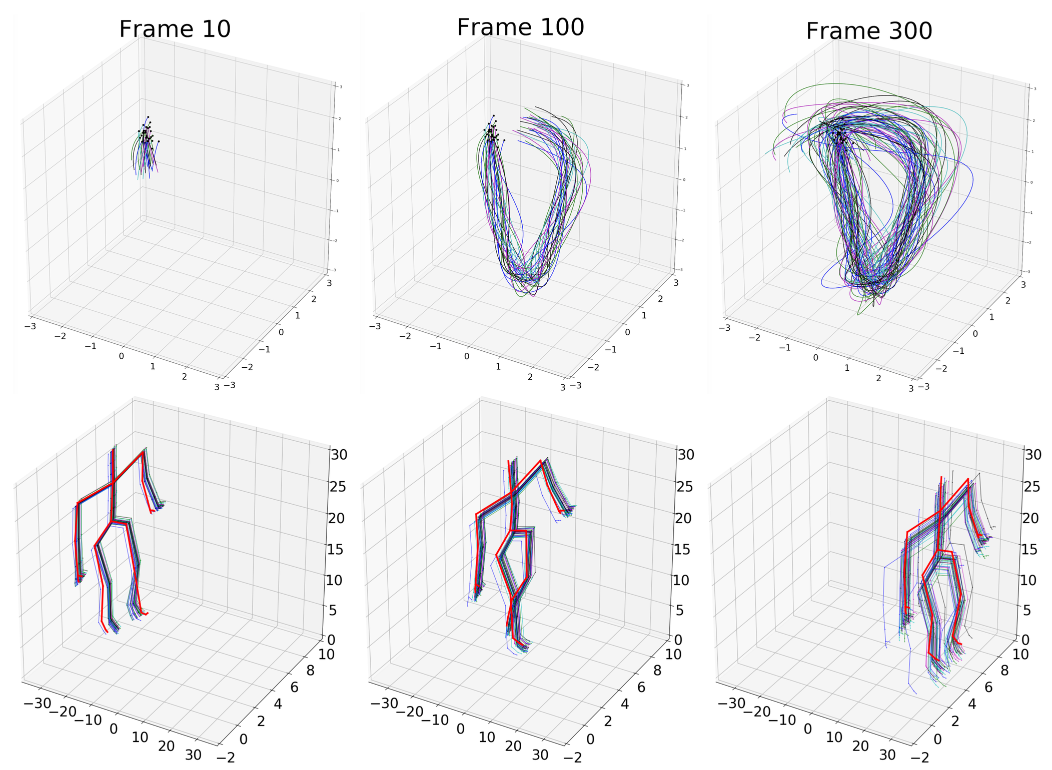
Bouncing Balls Experiment

Dataset: Sequences of three balls bouncing within a box (frames of size 32x32) [?].



CMU Walking Data Experiment

- 62-dim sensor measurements
- 12 training sequences
- 300 frames per sequence
- Predictions are conditioned on the first three frames
- ODE^{2}VAE outperforms neural ODEs (MSEs: 8.09 vs 22.49)
- Top row: 30 latent trajectory samples from learned model
- Bottom: Corresponding reconstructions



References

[1] Kingma, D., and Welling, M. "Auto-encoding variational bayes." arXiv:1312.6114 (2013).

[2] Chen, Tian Qi, et al. "Neural ordinary differential equations." NeurIPS, 2018.

[3] Casale, Francesco Paolo, et al. "Gaussian process prior variational autoencoders." NeurIPS, 2018.

[4] Hsieh, Jun-Ting, et al. "Learning to Decompose and Disentangle Representations for Video Prediction." NeurIPS, 2018.

Code and poster available at:

Contact: Çağatay Yıldız
cagatay.yildiz@aalto.fi