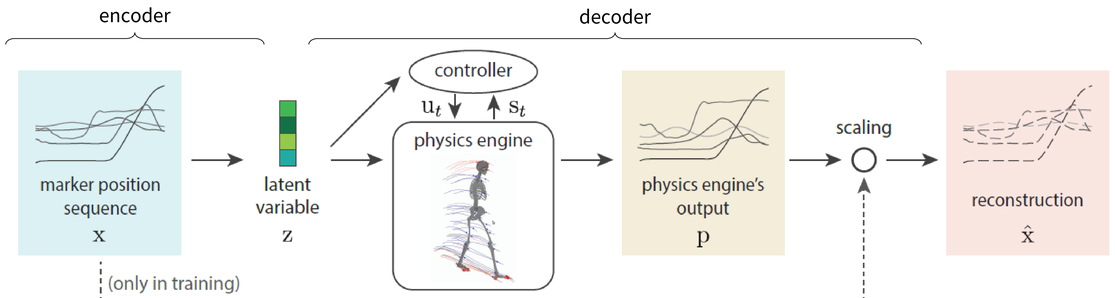
Variational Autoencoder with Differentiable Physics Engine for Human Gait Analysis and Synthesis

Naoya Takeishi, Alexandros Kalousis (University of Applied Sciences and Arts Western Switzerland, HES-SO)

We address the task of learning generative models of human gait. As gait motion always follows the physical laws, a generative model should also produce outputs that comply with the physical laws, particularly rigid body dynamics with contact and friction. We propose a variational autoencoder (VAE) combined with a differentiable physics engine, which outputs physically plausible signals by construction. It is also equipped with a policy network conditioned on each gait sample. Training is done as ordinary VAEs.

VAE with differentiable physics engine in decoder



Data of our interest is measurements of human gait (x) as sequences of position of markers attached to body taken by motion capture system. We would like to learn a generative model of human gait for analyzing and synthesizing gait patterns. We particularly use VAEs for dealing with new gait measurements (e.g., from new subjects, from subjects after surgery, etc.) in an amortized manner.

Encoder is with neural networks as with ordinary VAEs. It gives latent variable (z) for each sequence.

Decoder comprises a physics engine that simulates rigid body dynamics with contacts.

- *Agent* is an articulated rigid body that mimics human's skeleton with 21 DoF
- *Controller* is a neural network (policy network) that gives input (joint torque) to agent

Decoder outputs simulated trajectory of positions of markers (p) attached to the agent.

After **scaling** of the simulated trajectory p to the scale of original x, reconstruction ($\hat{x}$) is finally obtained. Learning is done by maximizing ELBO with such reconstruction.

Notably, we use *differentiable physics engine*, with which training is performed as ordinary VAEs with stochastic gradient descent. We used `nimblephysics`[1] [Werling+, arXiv:2103.16021].

[1] https://nimblephysics.org/

Details of the model

Architecture

- Encoder is `TransformerEncoder`: $\mathrm{z} \sim \mathcal{N}(\mathrm{z}; \boldsymbol{\mu}_{\mathrm{encoder}}(\mathrm{x}), \sigma^2_{\mathrm{encoder}}(\mathrm{x})\boldsymbol{I})$
- Initial condition of simulator is determined by latent variable: $\mathrm{s}_0 = \boldsymbol{f}_{\mathrm{initializer}}(\mathrm{z})$
- In simulator, at each timestep t:
  - Input (joint torque) is computed by MLP that takes current state and latent variable:
    $$\mathrm{u}_t = \boldsymbol{f}_{\mathrm{controller}}(\mathrm{s}_t, \mathrm{PositionalEncoding}(\mathrm{z}; t))$$
  - State (floating base attitude & joint angles) at next timestep is computed by the engine:
    $$\mathrm{s}_{t+1} = \mathrm{TemporalTransition}(\mathrm{s}_t, \mathrm{u}_t)$$
- Simulated marker position $\{p_t\}$ is computed from $\{s_t\}$ via forward kinematics
- Linear scaling is applied to $\{p_t\}$: $\mathbb{E}[\hat{\mathrm{x}}_{t,i,j}] = \alpha_i(\mathrm{p}_{t,i,j} - \mathrm{p}_{t,i,\mathrm{fb}}) + \beta_i \mathrm{p}_{t,i,\mathrm{fb}} + \gamma_i$
  where the coefficients are the solutions (by differentiable solver) of the following convex problem:
  $$\min_{\alpha_i,\beta_i,\gamma_i} \sum_{t=1}^{\ell}\sum_{j=1}^{m} |\mathbb{E}[\hat{\mathrm{x}}_{t,i,j}] - \mathrm{x}_{t,i,j}|^2 \quad \text{s.t.} \quad \alpha_i \in [\alpha_{\mathrm{lb}}, \alpha_{\mathrm{ub}}], \beta_i \in [\beta_{\mathrm{lb}}, \beta_{\mathrm{ub}}], \gamma_i \in [\gamma_{\mathrm{lb}}, \gamma_{\mathrm{ub}}]$$

Heuristics for stable training

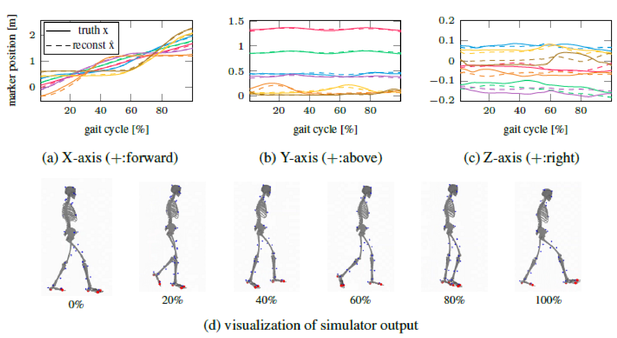
- Use not only x (marker position) but also numerically computed $\dot{x}$ (marker velocity) as data/recon
- Regularization terms
  - penalizing violation of joint angle constraints to objective
  - penalizing large power (torque × angular velocity)
  - penalizing large temporal difference of torque
- (IMPORTANT) Start training with short sequences and then feed longer sequences gradually

Preliminary experiments

We used public dataset [Lencioni+, Sci. Dat. 6(1):309, 2019]; 328 sequences of length 500 for training.

Reconstruction

Average test RMSE was 1.94 ± 0.36 [cm].



Style transfer

With gait cadence as conditional variable, we made z disentangled with it and altered z in test time.