The Gaussian Process Prior VAE for Latent Dynamics from Pixels

Michael Pearce

Complexity Science, University of Warwick, UK

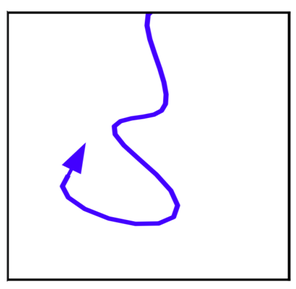
Research Question

- given videos of moving object $v_{1:T}$
- unsupervised learn latent $x_{1:T}, y_{1:T}$
- ~~graphical model prior~~ use a GP?
- Let's see on some toy data...

Generative Model

GP Prior

$$x_{1:T}, y_{1:T} \sim \mathcal{GP}\left(x_{1:T}, y_{1:T}\right)$$



NN+Bernoulli L'hood

$$v_{1:T} \sim \prod_{t=1}^{T} \mathcal{B}(v_t|p_\theta(x_t, y_t))$$

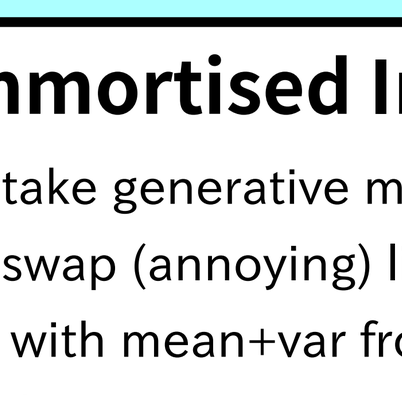
$$\mathbb{P}[v_{1:T}, x_{1:T}, y_{1:T}] = \mathcal{GP}\left(x_{1:T}, y_{1:T}\right) \prod_{t=1}^{T} \mathcal{B}(v_t|p_\theta(x_t, y_t))$$

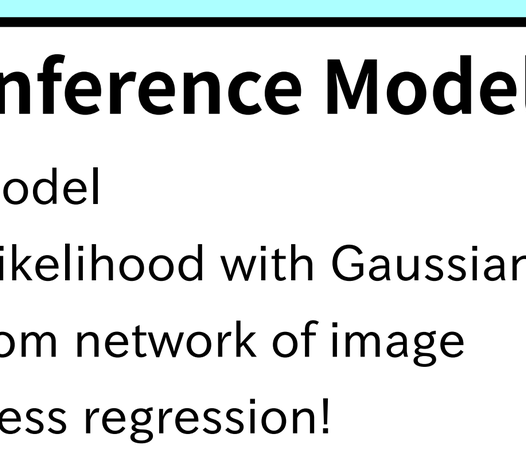
Ammortised Inference Model

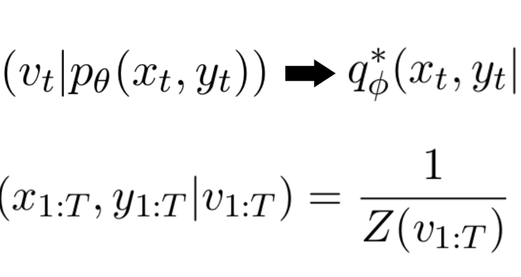
1. take generative model
2. swap (annoying) likelihood with Gaussian with mean+var from network of image
3. do Gaussian process regression!

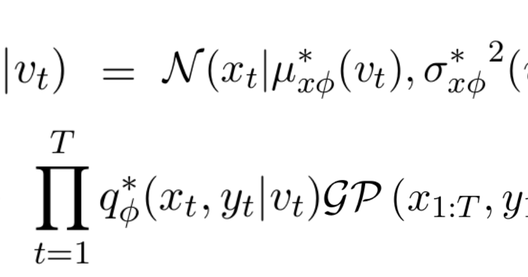
$$\mathcal{B}(v_t|p_\theta(x_t, y_t)) \Rightarrow q^*_\phi(x_t, y_t|v_t) = \mathcal{N}(x_t|\mu^*_{x\phi}(v_t), {\sigma^*_{x\phi}}^2(v_t))$$

$$q(x_{1:T}, y_{1:T}|v_{1:T}) = \frac{1}{Z(v_{1:T})} \prod_{t=1}^{T} q^*_\phi(x_t, y_t|v_t)\mathcal{GP}\left(x_{1:T}, y_{1:T}\right)$$



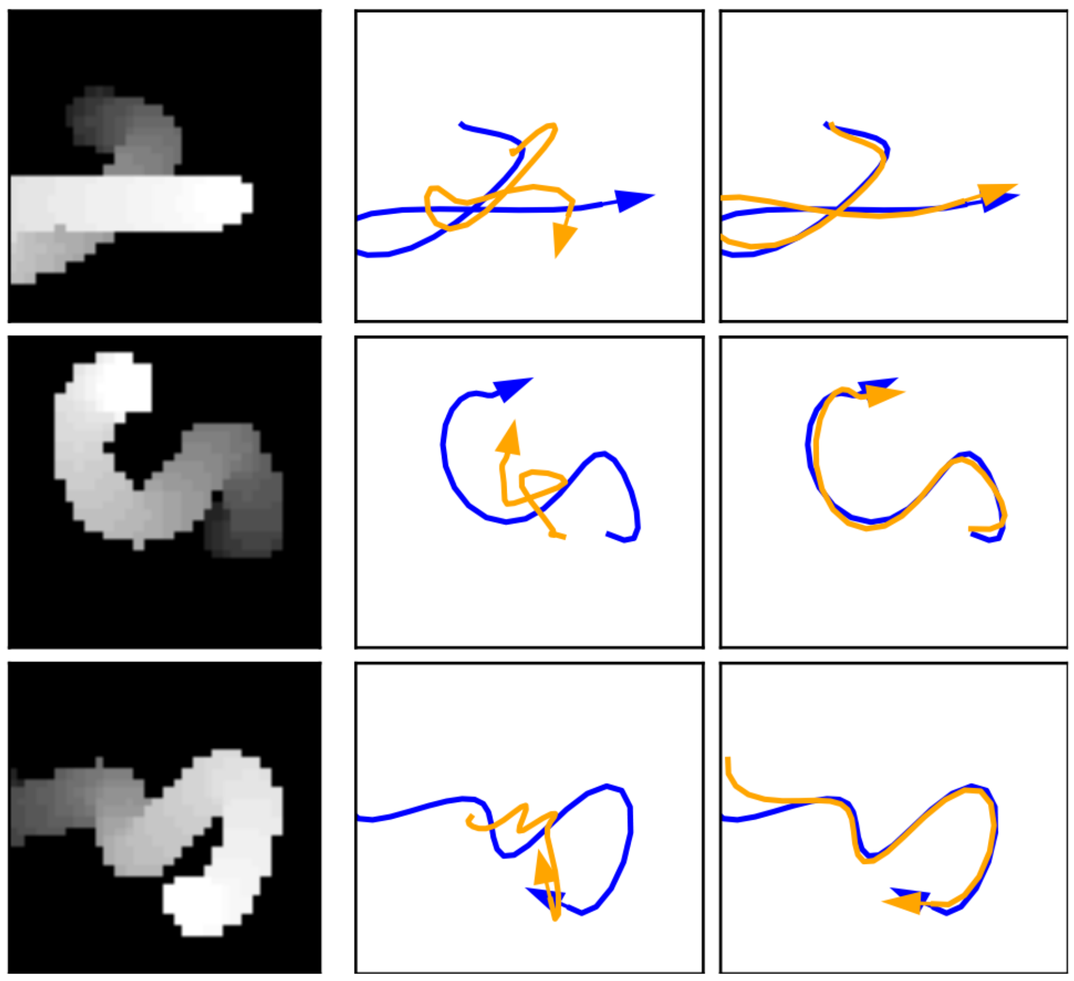




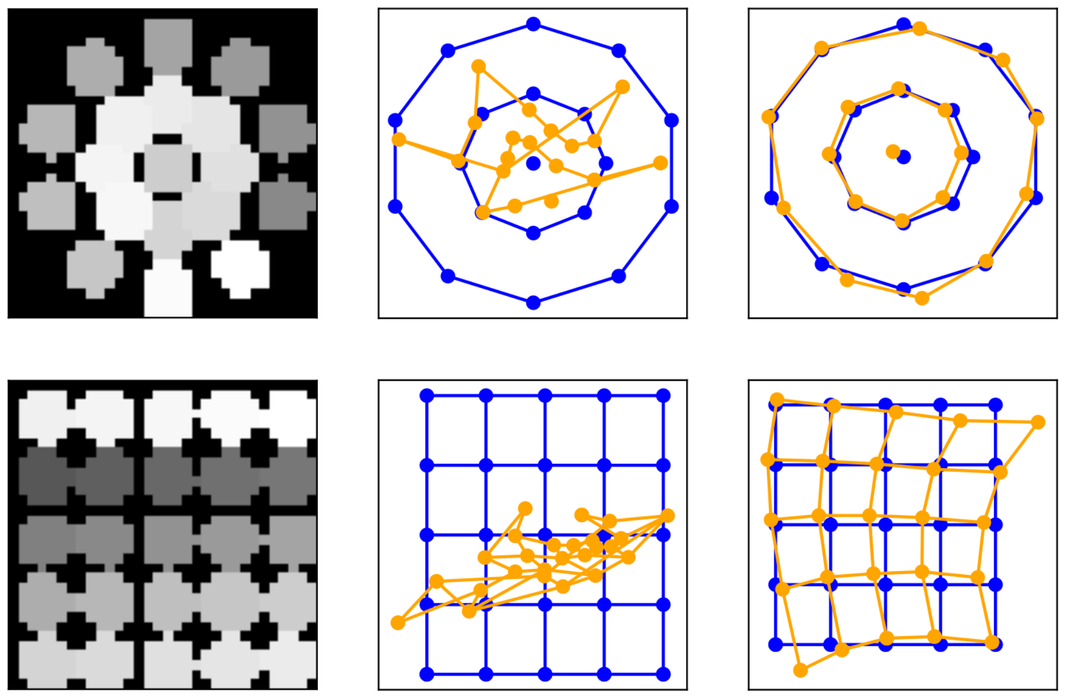


Experiments

$$\mathbb{E}_q \left[\underbrace{\sum_{t \in T} \log \mathcal{B}(v_t|p_\theta(x_t, y_t))}_{\text{recon (MC)}} \underbrace{- \log q^*_\phi(x_t, y_t|v_t)}_{\text{KL prior (analytic)}} \right] + \underbrace{\log Z(v_{1:T})}_{\text{GPR l'hood}}$$



How "Euclidean" is the latent space?



Many open questions

- model-mismatch: non GP-synthetic-data
- kernels: periodic/Brownian/Matern
- Bayesian forcasting/planning?
- reduce cubic computation of inference