Composing graphical models and neural networks for structured representations and fast inference

**Motivation**

- **TL;DR**: variational autoencoders + latent graphical models
- **modeling idea**: use PGM priors to organize the latent space, along with neural net observation models for flexible representations

**Inference idea**: use PGMs to synthesize information from recognition nets instead of making a single inference net do everything

**Probabilistic graphical models**
- structured representations
- priors and uncertainty
- data and computational efficiency within rigid model classes
- rigid assumptions may not fit
- feature engineering
- more flexible models can require slow top-down inference

**Deep neural networks**
- neural net “gpm”
- difficult parameterization
- can require lots of data
- flexible, high capacity
- feature learning
- recognition networks for fast bottom-up inference

**Automatically learn representations in which structured PGMs fit well**

**Learning to parse mouse behavior from depth video**

**Inference**

- **natural gradient SVI**
  - \( p(x) = \arg \max_{q(x)} \mathcal{L}(q(x) \mid q(x)) \)
  - optimal local factor
  - expensive for general obs.
  - explicit conjugate structure
  - arbitrary inference queries
  - natural gradients

- **variational autoencoders**
  - \( q(x) = \mathcal{N}(x \mid \mu(y; \phi), \Sigma(y; \phi)) \)
  - suboptimal local factor
  - fast for general obs.
  - \( \phi \) does all local inference
  - limited inference queries
  - no natural gradients

- **structured VAEs (this work)**
  - \( q(x) = \mathcal{N}(x \mid \mu(y; \phi), \Sigma(y; \phi)) \)
  - optimal given conj. evidence
  - fast for general obs.
  - explicit conjugate structure
  - arbitrary inference queries
  - some natural gradients

**Main idea**

learn to summarize complicated evidence with simple conjugate potentials (as in CRFs)

\[
\mathbb{E}_{q(z)} \log p(y_n \mid x_n, \gamma)
\]

\( \phi(x_n; y_n, \phi) \)

**Proposition (log evidence lower bound)**

Legend: dart, pause, rear

**SVAE**

fit a latent switching linear dynamical system (SLDS) and a neural network image model for observations

\[
\mathcal{L}(y, \eta_1, \eta_2) \triangleq \mathbb{E}_q(y', \eta_1, \eta_2) \left[ \log p(y', \eta_1, \eta_2) + \log \phi(y', \eta_1, \eta_2) \right] \\
\mathcal{L}(y, \eta_1, \eta_2) \triangleq \mathbb{E}_q(y', \eta_1, \eta_2) \left[ \log p(y', \eta_1, \eta_2) + \log \phi(y', \eta_1, \eta_2) \right]
\]