Summary of paper #3
Structured Inference Networks for Nonlinear State Space Models

This paper proposes a unified learning algorithm for a broad class of **Gaussian State Space Models (GSSMs)**. Authors introduce an inference procedure that scales easily to high dimensional data, both in the observed and the latent spaces without compromising the quality of inference and learning, compiling approximate (and where feasible, exact) inference into the parameters of a neural network.

the learning algorithm performs **stochastic gradient ascen**t on a **variational lower bound of the likelihood**. Instead of introducing variational parameters for each data point, they *compile* the inference procedure at the same time as learning the generative model.

More specifically, they introduce a new family of *structured inference networks*, parameterized by recurrent neural networks, and evaluate their effectiveness in three scenarios: (1) when the generative model is known and fixed, (2) in parameter estimation when the functional form of the model is known and (3) for learning deep Markov models. By looking at the structure of the true posterior, they show both theoretically and empirically that inference for a latent state should be performed using information *from its future.*
This approach may easily be adapted to learning more general generative models, for example, models with edges from observations to latent states. They also show that for generative models, the posterior distribution at any time step is a function of all future (and past) observations. Additionally, they systematically evaluate the impact of the different variational approximations on learning.
Finally, they learn a DMM on a polyphonic music dataset and on a dataset of electronic health records (a complex high dimensional setting with missing data). they use the model learned on health records to ask queries such as "what would have happened to patients had they not received treatment", and show that their model correctly identifies the way certain medications affect a patient's health.

What are the benefits of this approach?
1. Using RNN in inference network → it is possible to continue to increase their capacity and condition on different modalities → relevant to approximate posterior inference (without worrying about overfitting the data).
2. The semantics of the DMM render it easy to marginalize out unobserved data.

Limitations?

1. They didn't consider the robustness of the model against parameter uncertainties.