Approximate Probabilistic Inference with Composed Flows

Jay Whang, Erik M. Lindgren, Alexandros G. Dimakis

The University of Texas at Austin

NEURAL INFORMATION PROCESSING SYSTEMS

Why Inference on a Flow?

Motivation

- Generative models are becoming increasingly expensive to train. So we would like to **repurpose** a single pre-trained model to perform multiple downstream tasks.
- One way to reuse the same generative model for multiple tasks is to perform **probabilistic inference**.
- Normalizing flow models are **flexible**, offering fast likelihood evaluation, sampling, and inversion.

Problem

- Given a joint distribution defined by a flow, how can we perform **conditional inference** for a given observation while **retaining the computational flexibility** of a flow model?
- In other words, can we learn a flow model that approximates the conditional $p(\boldsymbol{x}_2 \mid \boldsymbol{x}_1 = \boldsymbol{x}_1^*)$ for the given observation $\boldsymbol{x}_1^*$, from the base model $p(\boldsymbol{x}) = p(\boldsymbol{x}_1, \boldsymbol{x}_2)$?

Hardness of Inference

- Conditional inference in a general Bayesian Belief Network is known to be NP-hard. Perhaps the computational flexibility of a flow model makes inference tractable? Unfortunately not.
- We show that sampling from the exact conditional distribution is hard for a large family of existing flow architectures. Moreover, even **approximate** sampling is hard.
- This motivates the use of **approximate conditioning**, where we perform Gaussian smoothing on the observed variable to allow the given observation to be matched with some error.
- Thus, the goal is to learn $p(\boldsymbol{x}_2 \mid \tilde{\boldsymbol{x}}_1 = \boldsymbol{x}_1^*)$ as a flow model where $\tilde{\boldsymbol{x}}_1 \sim \mathcal{N}(\boldsymbol{x}_1, \sigma^2 \boldsymbol{I})$ is the smoothed version of $\boldsymbol{x}_1$.

Our Approach: Pre-generator

Learning to Construct Noise with a Pre-generator

- We propose to learn a flow model defined by the invertible mapping $\hat{f} : \boldsymbol{\epsilon} \mapsto \boldsymbol{z}$ such that its **composition** with the given model produces approximate conditional samples:

$$\boldsymbol{\epsilon} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}) \Rightarrow (f \circ \hat{f})(\boldsymbol{\epsilon}) \sim p(\boldsymbol{x} \mid \tilde{\boldsymbol{x}}_1 = \boldsymbol{x}_1^*)$$

  where $f : \boldsymbol{z} \mapsto \boldsymbol{x}$ is the invertible mapping for the joint model.
- Because the composition $f \circ \hat{f}$ is **invertible**, this approximate posterior (which we denote $q(\boldsymbol{x})$) is itself a flow model and allows for exact likelihood evaluation, fast sampling, and inversion.

Training the Pre-generator

Modified VI Objective

- Because we have access to samples and likelihood only through the base model, we cannot directly compute the marginal posterior:

$$p(\boldsymbol{x}_2 \mid \tilde{\boldsymbol{x}}_1 = \boldsymbol{x}_1^*) = \int_{\boldsymbol{x}_1} p(\boldsymbol{x} \mid \tilde{\boldsymbol{x}}_1 = \boldsymbol{x}_1^*) \mathrm{d}\boldsymbol{x}_1$$

- We instead optimize the **joint variational posterior** via stochastic variational inference by minimizing the KL $D_{\mathrm{KL}}\left(q(\boldsymbol{x}) \| p(\boldsymbol{x} \mid \tilde{\boldsymbol{x}}_1 = \boldsymbol{x}_1^*)\right)$, which is an **upper bound** to the intractable marginal KL $D_{\mathrm{KL}}\left(q(\boldsymbol{x}_2) \| p(\boldsymbol{x}_2 \mid \tilde{\boldsymbol{x}}_1 = \boldsymbol{x}_1^*)\right)$.

Generalization to Transformed Observations

- The VI formulation allows for an easy generalization to conditioning under a differentiable transformation T, where we observe some $\boldsymbol{y}^*$ in the range of T. The corresponding objective is ($\approx$ denotes Gaussian smoothing):

$$D_{\mathrm{KL}}\left(q(\boldsymbol{x}) \| p(\boldsymbol{x} \mid T(\boldsymbol{x}) \approx \boldsymbol{y}^*)\right)$$

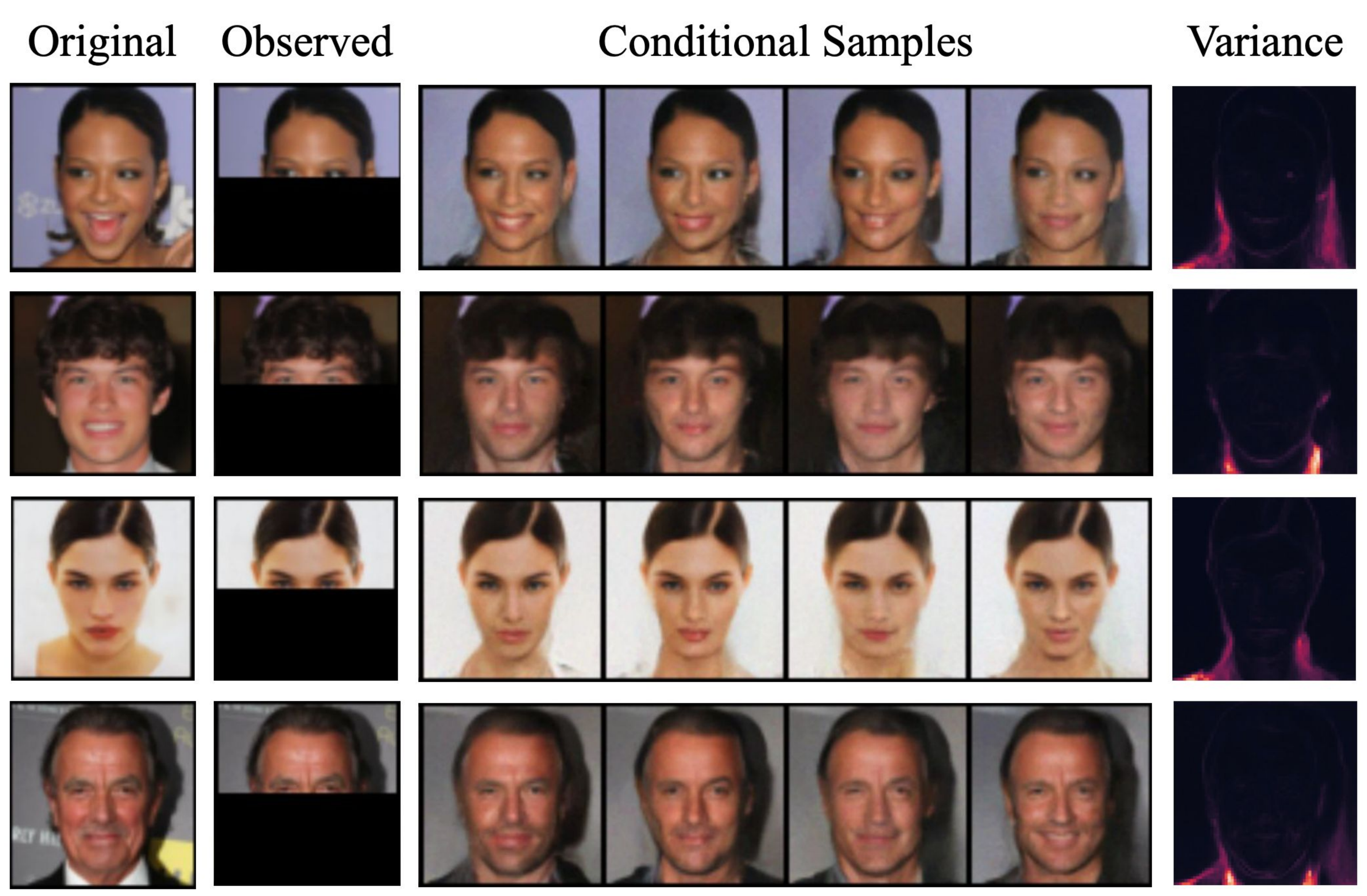
  which further simplifies to

$$\mathcal{L}(q) = D_{\mathrm{KL}}\left(q(\boldsymbol{z}) \| p(\boldsymbol{z})\right) + \mathbb{E}_{\boldsymbol{z} \sim q}\left[\frac{1}{2\sigma^2} \| T(f(\boldsymbol{z})) - \boldsymbol{y}^* \|_2^2\right]$$

Experimental Results

Image Completion on CelebA-HQ

- Conditional Completions



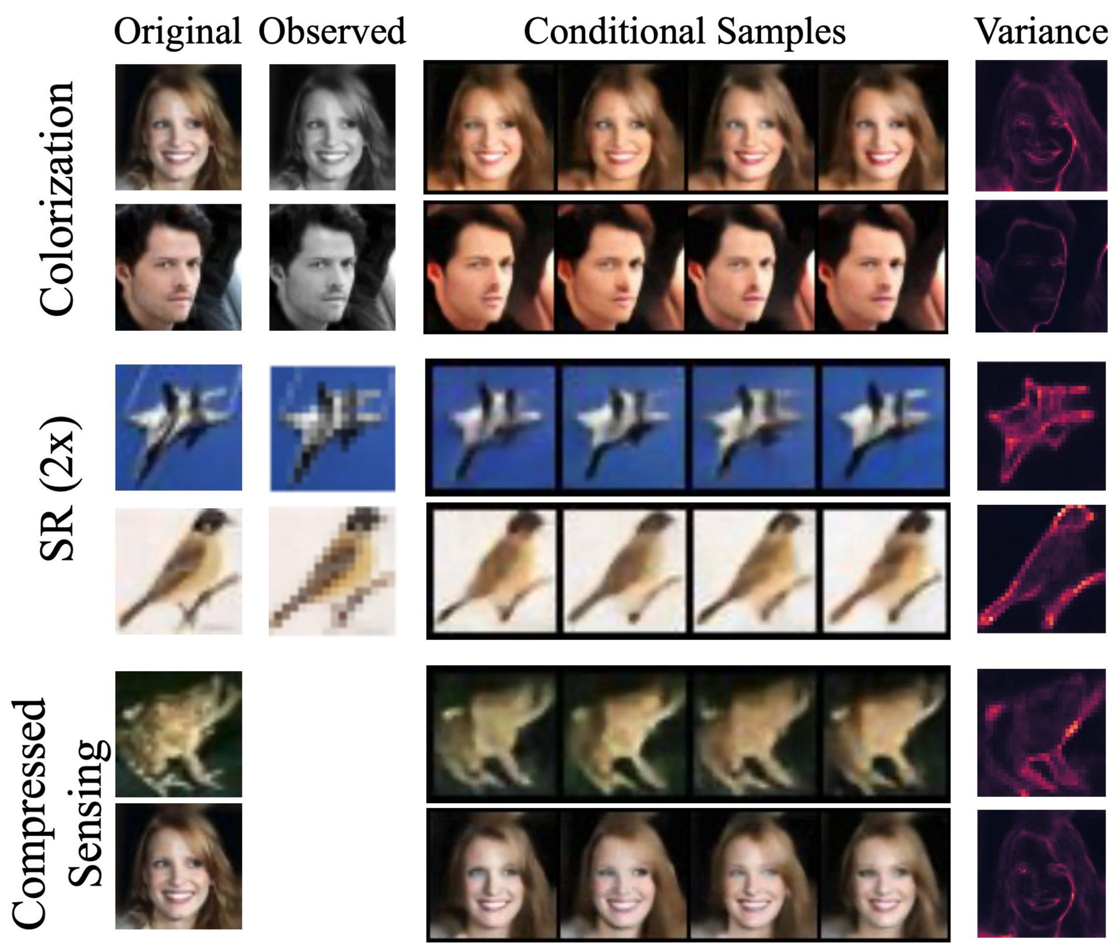
Experimental Results

- Sample quality compared to baselines (variational inference in image space, Langenvin dynamics and PL-MCMC)

	MNIST			CIFAR-10 (5-bit)				CelebA-HQ (5-bit)		
	FID	MSE	LPIPS	FID	IS↑	MSE	LPIPS	FID	MSE	LPIPS
Ours	**4.11**	**21.67**	**0.074**	**41.14**	**7.189**	9.71	**0.176**	33.61	**223.06**	**0.208**
Langevin	14.34	36.51	0.135	47.53	6.732	**9.31**	0.201	**30.33**	323.47	0.229
Ambient VI	114.59	65.56	0.290	84.78	5.156	16.74	0.296	289.64	1060.66	0.587
PL-MCMC	21.20	59.89	0.190	N/A				N/A		

- Class-conditional Sampling from an unconditional model

- Various inverse problems



Conclusion

- Flow-based variational inference is stable to train and has many computational advantages, without a large degradation in the sample quality compared to simple MCMC baselines.
- Future directions include amortization as well as generalization to non-invertible generators.