Flow++: Improving Flow-Based Generative Models with Variational Dequantization and Architecture Design

Jonathan Ho*, Xi Chen*, Aravind Srinivas, Yan Duan, Pieter Abbeel

Overview

- Goal: likelihood-based model with
  - Fast sampling and training
  - Good samples and density estimation performance
- Our strategy: improve flow models
  - Uniform dequantization -> variational dequantization
  - Affine coupling -> mixture of logistics coupling
  - Convolutions -> convolutions + self-attention

Continuous flows for discrete data

- A problem arises when fitting continuous density models to discrete data: degeneracy
  - When the data are 3-bit pixel values, $\mathbf{x} \in \{0, 1, 2, \ldots, 255\}$
  - What density does a model assign to values between bins like 0.4, 0.42…?
- Correct semantics: we want the integral of probability density within a discrete interval to approximate discrete probability mass

$$P_{\text{model}}(\mathbf{x}) := \int_{[0,1)^D} p_{\text{model}}(\mathbf{x} + \mathbf{u}) \, d\mathbf{u}$$

Continuous flows for discrete data

- Solution: **Dequantization**. Add noise to data.
  - $\mathbf{x} \in \{0, 1, 2, \ldots, 255\}$
  - We draw noise u uniformly from $[0, 1)^D$

$$\begin{aligned}
\mathbb{E}_{\mathbf{y}\sim p_{\text{data}}}[\log p_{\text{model}}(\mathbf{y})] &= \sum_{\mathbf{x}} P_{\text{data}}(\mathbf{x}) \int_{[0,1)^D} \log p_{\text{model}}(\mathbf{x}+\mathbf{u})\, d\mathbf{u} \\
&\leq \sum_{\mathbf{x}} P_{\text{data}}(\mathbf{x}) \log \int_{[0,1)^D} p_{\text{model}}(\mathbf{x}+\mathbf{u})\, d\mathbf{u} \\
&= \mathbb{E}_{\mathbf{x}\sim P_{\text{data}}}[\log P_{\text{model}}(\mathbf{x})]
\end{aligned}$$

[Theis, Oord, Bethge, 2016]

Variational Dequantization

- **Variational Dequantization**. Add a learnable noise q to data.

$$\begin{aligned}
\mathbb{E}_{\mathbf{x}\sim P_{\text{data}}}\left[\log P_{\text{model}}(\mathbf{x})\right] &= \mathbb{E}_{\mathbf{x}\sim P_{\text{data}}}\left[\log \int_{[0,1)^D} q(\mathbf{u}|\mathbf{x}) \frac{p_{\text{model}}(\mathbf{x}+\mathbf{u})}{q(\mathbf{u}|\mathbf{x})}\, d\mathbf{u}\right] \\
&\geq \mathbb{E}_{\mathbf{x}\sim P_{\text{data}}}\left[\int_{[0,1)^D} q(\mathbf{u}|\mathbf{x}) \log \frac{p_{\text{model}}(\mathbf{x}+\mathbf{u})}{q(\mathbf{u}|\mathbf{x})}\, d\mathbf{u}\right] \\
&= \mathbb{E}_{\mathbf{x}\sim P_{\text{data}}}\mathbb{E}_{\mathbf{u}\sim q(\cdot|\mathbf{x})}\left[\log \frac{p_{\text{model}}(\mathbf{x}+\mathbf{u})}{q(\mathbf{u}|\mathbf{x})}\right]
\end{aligned}$$

[Ho et al., 2019]

Coupling layers

RealNVP

$$\mathbf{y}_1 = \mathbf{x}_1$$

$$\mathbf{y}_2 = \mathbf{x}_2 \cdot \exp(\mathbf{a}_\theta(\mathbf{x}_1)) + \mathbf{b}_\theta(\mathbf{x}_1)$$

convolutions

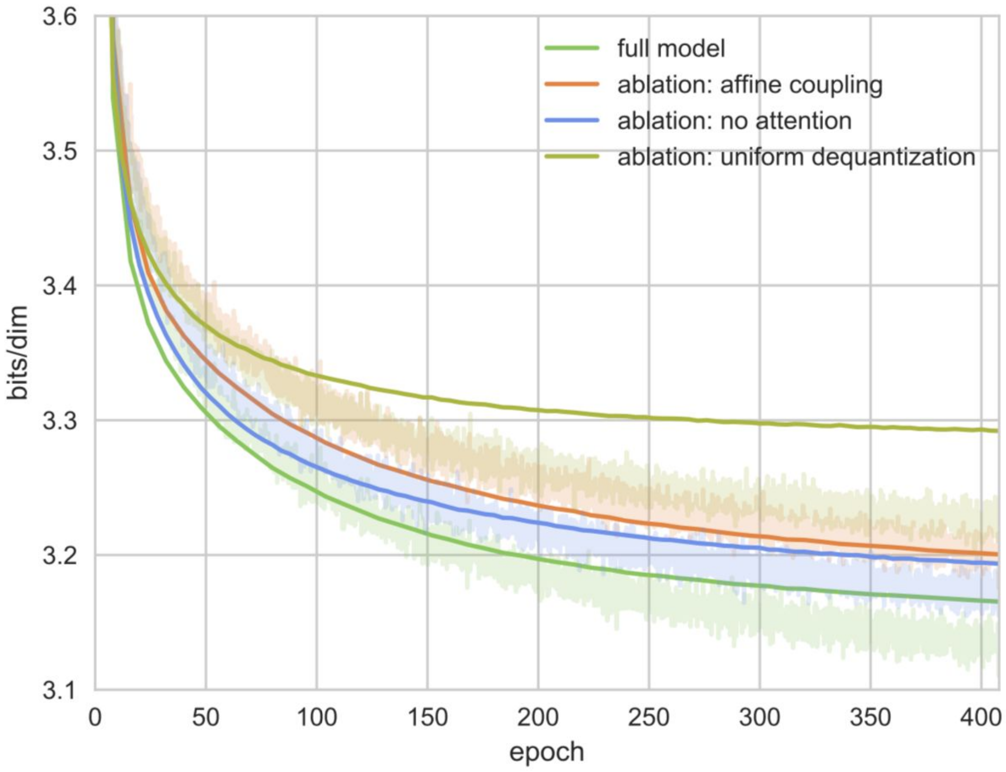
Ours: logistic mixture CDF

$$\mathbf{y}_1 = \mathbf{x}_1$$

$$\mathbf{y}_2 = \sigma^{-1}\left(\mathrm{MixLogCDF}(\mathbf{x}_2; \pi_\theta(\mathbf{x}_1), \mu_\theta(\mathbf{x}_1), \mathbf{s}_\theta(\mathbf{x}_1))\right) \cdot \exp(\mathbf{a}_\theta(\mathbf{x}_1)) + \mathbf{b}_\theta(\mathbf{x}_1)$$

convolutions & self-attention

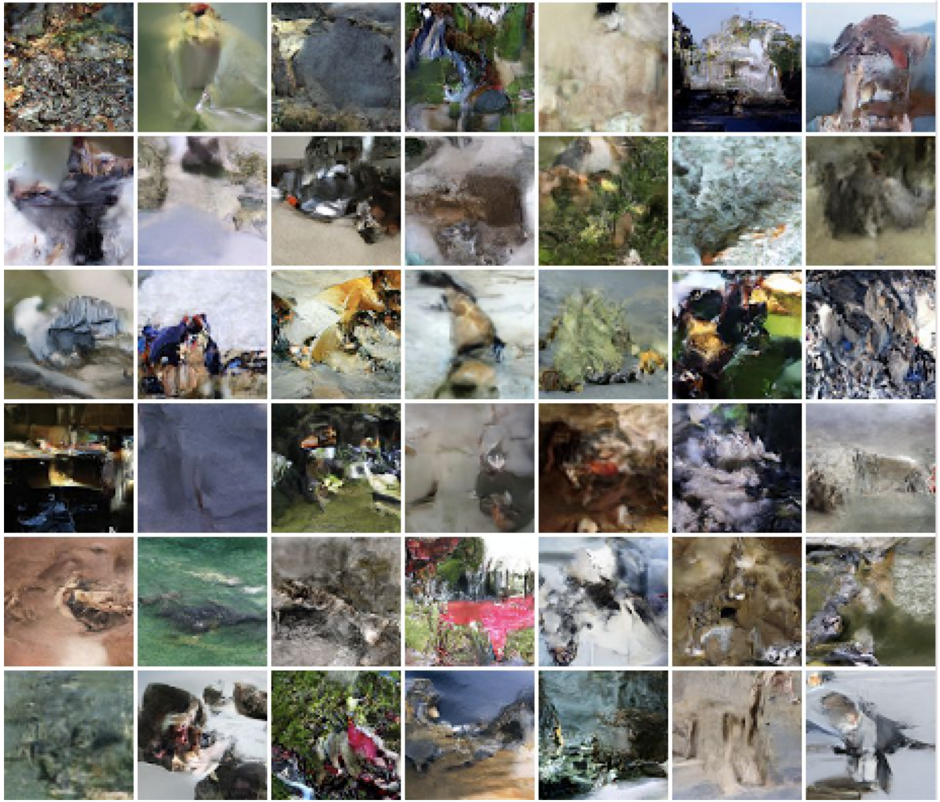
Ablation on CIFAR



Results

Model family	Model	CIFAR10	ImageNet 32x32	ImageNet 64x64
Non-autoregressive	RealNVP (Dinh et al., 2016)	3.49	4.28	–
	Glow (Kingma & Dhariwal, 2018)	3.35	4.09	3.81
	IAF-VAE (Kingma et al., 2016)	3.11	–	–
	Flow++ (ours)	**3.08**	**3.86**	**3.69**
Autoregressive	Multiscale PixelCNN (Reed et al., 2017)	–	3.95	3.70
	PixelCNN (van den Oord et al., 2016b)	3.14	–	–
	PixelRNN (van den Oord et al., 2016b)	3.00	3.86	3.63
	Gated PixelCNN (van den Oord et al., 2016c)	3.03	3.83	3.57
	PixelCNN++ (Salimans et al., 2017)	2.92	–	–
	Image Transformer (Parmar et al., 2018)	2.90	3.77	–
	PixelSNAIL (Chen et al., 2017)	2.85	3.80	3.52

Samples (CIFAR10, ImageNet 64x64)



Samples (CelebA 5-bit)

- Slides adapted from Berkeley CS294-158 Deep Unsupervised Learning class: https://sites.google.com/view/berkeley-cs294-158-sp19/home

    - Want to learn more about foundation of Deep Generative Models & Self-Supervised learning methods?
    - All lecture videos are available on youtube, featuring guest speakers: Ilya Sutskever, Alyosha Efros, Alec Radford, Aaron van den Oord