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, $x \in \{0, 1, 2, \ldots, 255\}$
  - 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

$$P_{model}(x) := \int_{[0,1)^D} p_{model}(x + u) \, du$$
Continuous flows for discrete data

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

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

[Theis, Oord, Bethge, 2016]
Variational Dequantization. Add a learnable noise $q$ to data.

\[
\mathbb{E}_{x \sim p_{\text{data}}} \left[ \log p_{\text{model}}(x) \right] = \mathbb{E}_{x \sim p_{\text{data}}} \left[ \log \int_{[0,1]^D} q(u|x) \frac{p_{\text{model}}(x + u)}{q(u|x)} \, du \right]
\]

\[
\geq \mathbb{E}_{x \sim p_{\text{data}}} \left[ \int_{[0,1]^D} q(u|x) \log \frac{p_{\text{model}}(x + u)}{q(u|x)} \, du \right]
\]

\[
= \mathbb{E}_{x \sim p_{\text{data}}} \mathbb{E}_{u \sim q(.|x)} \left[ \log \frac{p_{\text{model}}(x + u)}{q(u|x)} \right]
\]

[Ho et al., 2019]
Coupling layers

RealNVP

\[ y_1 = x_1 \]
\[ y_2 = x_2 \cdot \exp(a_\theta(x_1)) + b_\theta(x_1) \]

convolutions

Ours: logistic mixture CDF

\[ y_1 = x_1 \]
\[ y_2 = \sigma^{-1} \left( \text{MixLogCDF}(x_2; \pi_\theta(x_1), \mu_\theta(x_1), s_\theta(x_1)) \right) \cdot \exp(a_\theta(x_1)) + b_\theta(x_1) \]

convolutions & self-attention
Ablation on CIFAR
## Results

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