Adaptive Density Estimation for Generative Models

Thomas Lucas
Konstantin Shmelkov*
Karteek Alahari
Cordelia Schmid
Jakob Verbeek

* Now at Huawei
Goal

Given samples from target distribution $p^*$, train a model $p_\theta$ to match $p^*$
Generative modelling

Goal

Given samples from target distribution $p^*$, train a model $p_\theta$ to match $p^*$

- **Maximum likelihood**: Eval. training points under the model
Goal

Given samples from target distribution $p^*$, train a model $p_\theta$ to match $p^*$

- **Maximum likelihood**: Eval. training points under the model

- **Adversarial training**¹: Eval. samples under (approximation of) $p^*$

¹ Ian Goodfellow et al. (2014). “Generative adversarial nets”. In: NIPS.
Schematic illustration

Data

Model
Maximum likelihood

- MLE covers full support of distribution
- Produces unrealistic samples

Model

Data

data samples
Maximum likelihood covers the full support of the distribution, but it can also produce unrealistic samples. Over-generalization leads to data samples falling outside the model's predicted distribution.
Maximum likelihood

- MLE covers **full support** of distribution
- Produces **unrealistic samples**
Adversarial training

Mode-dropping

Consequences

• Production of high quality samples
• Parts of the support are dropped
Adversarial training

Consequences

- Production of **high quality** samples
- Parts of the support are **dropped**
Hybrid training approach

Goal

- Explicitly optimize both dataset coverage and sample quality
Hybrid training approach

Goal

- Explicitly optimize both dataset **coverage** and sample **quality**
- Discriminator can be seen as a **learnable** inductive bias
Hybrid training approach

Goal

- Explicitly optimize both dataset coverage and sample quality
- Discriminator can be seen as a learnable inductive bias
- Retain valid likelihood to evaluate support coverage
Hybrid training approach

Goal

- Explicitly optimize both dataset coverage and sample quality
- Discriminator can be seen as a learnable inductive bias
- Retain valid likelihood to evaluate support coverage

Challenges

- Tradeoff between the two objectives: need more flexibility
Hybrid training approach

Goal

- Explicitly optimize both dataset **coverage** and sample **quality**
- Discriminator can be seen as a **learnable** inductive bias
- Retain valid **likelihood** to evaluate support coverage

Challenges

- **Tradeoff** between the two objectives: need more flexibility
- Limiting **parametric assumptions** required for tractable MLE, e.g. Gaussianity, conditional independence
Hybrid training approach

**Goal**

- Explicitly optimize both dataset **coverage** and sample **quality**
- Discriminator can be seen as a **learnable** inductive bias
- Retain valid **likelihood** to evaluate support coverage

**Challenges**

- **Tradeoff** between the two objectives: need more flexibility
- Limiting **parametric assumptions** required for tractable MLE, e.g. Gaussianity, conditional independence
- Often no likelihood in pixel space\(^2\)

---

\(^2\)A. Larsen et al. (2016). “Autoencoding beyond pixels using a learned similarity metric”. In: ICML.
Conditional independence
Conditional independence

$$p(x|z) = \prod_i \mathcal{N}(x_i|\mu_\theta(z),\sigma_1)$$
Conditional independence

\[
p(x | z) = \prod_{i} \mathcal{N}(x_i | \mu_\theta(z), \sigma I_n)
\]
Going beyond conditional independence

Avoiding strong parametric assumptions

- Lift reconstruction losses into a feature space
Going beyond conditional independence

Avoiding strong parametric assumptions

- Lift reconstruction losses into a feature space
- Deep invertible models: valid density in image space
Avoiding strong parametric assumptions

- Lift reconstruction losses into a feature space
- Deep invertible models: valid density in image space
- Retain fast sampling for adversarial training
Maximum likelihood estimation with feature targets

\[ L_{\text{adv}}(p_{\theta}, \psi) = -\mathbb{E}_{p_{\theta}(z)} \left[ \ln D(f^{-1}(\psi(\mu_{\theta}(z)))) \right] \]
Maximum likelihood estimation with feature targets

Amortized Variational inference in feature space:

\[ L_{\theta,\phi,\psi}(x) = -\mathbb{E}_{q_{\phi}(z|x)} \left[ \ln(p_{\theta}(f_{\psi}(x)|z)) \right] + D_{KL}(q_{\phi}(z|x)\|p_{\theta}(z)) - \ln \left| \det \frac{\partial f_{\psi}}{\partial x} \right| \]

Evidence lower bound in feature space
Maximum likelihood estimation with feature targets

Amortized Variational inference in feature space:

\[ \mathcal{L}_{\theta,\phi,\psi}(x) = -\mathbb{E}_{q_{\phi}(z|x)} \left[ \ln(p_{\theta}(f_{\psi}(x)|z)) \right] + D_{KL}(q_{\phi}(z|x)||p_{\theta}(z)) - \ln \left| \det \frac{\partial f_{\psi}}{\partial x} \right| \]

Change of variable
Amortized Variational inference in feature space:

\[
L_{\theta,\phi,\psi}(x) = -\mathbb{E}_{q_{\phi}(z|x)} \left[ \ln (p_{\theta}(f_{\psi}(x)|z)) \right] + D_{KL}(q_{\phi}(z|x) || p_{\theta}(z)) - \ln \left| \det \frac{\partial f_{\psi}}{\partial x} \right|
\]
Maximum likelihood estimation with feature targets

Amortized Variational inference in feature space:

\[
\mathcal{L}_{\theta, \phi, \psi}(x) = -\mathbb{E}_{q_{\phi}(z|x)} \left[ \ln(p_{\theta}(f_{\psi}(x)|z)) \right] + D_{KL}(q_{\phi}(z|x)||p_{\theta}(z)) - \ln \left| \det \frac{\partial f_{\psi}}{\partial x} \right|
\]

Adversarial training with Adaptive Density Estimation:

\[
\mathcal{L}_{\text{adv}}(p_{\theta}, \psi) = -\mathbb{E}_{p_{\theta}(z)} \left[ \ln \left( \frac{D(f_{\psi}^{-1}(\mu_{\theta}(z)))}{1 - D(f_{\psi}^{-1}(\mu_{\theta}(z)))} \right) \right]
\]

Adv. update using log ratio loss
Experiments on CIFAR10

<table>
<thead>
<tr>
<th>Model</th>
<th>IS</th>
<th>FID</th>
</tr>
</thead>
<tbody>
<tr>
<td>GAN</td>
<td></td>
<td>7.9</td>
</tr>
<tr>
<td>WGAN-GP</td>
<td>8.0</td>
<td></td>
</tr>
<tr>
<td>SNGAN</td>
<td>7.1</td>
<td>2.9</td>
</tr>
<tr>
<td>SNGAN (R,H)</td>
<td>8.2</td>
<td>2.7</td>
</tr>
<tr>
<td>MLE</td>
<td>3.1</td>
<td>3.8</td>
</tr>
<tr>
<td>VAE-IAF</td>
<td></td>
<td>3.7</td>
</tr>
<tr>
<td>NVP</td>
<td>3.5</td>
<td>4.5</td>
</tr>
<tr>
<td>Hybrid</td>
<td>3.8</td>
<td>8.2</td>
</tr>
<tr>
<td>Ours (v1)</td>
<td>3.5</td>
<td>6.9</td>
</tr>
<tr>
<td>Ours (v2)</td>
<td>3.5</td>
<td>6.9</td>
</tr>
<tr>
<td>FlowGan</td>
<td>4.2</td>
<td>3.9</td>
</tr>
</tbody>
</table>

*Note: Columns not aligned properly in table.*
### Experiments on CIFAR10

#### Samples Real images

<table>
<thead>
<tr>
<th>Model</th>
<th>BPD ↓</th>
<th>IS ↑</th>
<th>FID ↓</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>GAN</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WGAN-GP</td>
<td>7.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SNGAN</td>
<td>7.4</td>
<td>29.3</td>
<td></td>
</tr>
<tr>
<td>SNGAN(_{R,H})</td>
<td>8.2</td>
<td>21.7</td>
<td></td>
</tr>
<tr>
<td><strong>MLE</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VAE-IAF</td>
<td>3.1†</td>
<td>73.5†</td>
<td></td>
</tr>
<tr>
<td>NVP</td>
<td>3.5†</td>
<td>56.8†</td>
<td></td>
</tr>
<tr>
<td><strong>Hybrid</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ours (v1)</td>
<td>3.8</td>
<td>8.2</td>
<td>17.2</td>
</tr>
<tr>
<td>Ours (v2)</td>
<td>3.5</td>
<td>6.9</td>
<td>28.9</td>
</tr>
<tr>
<td>FlowGan</td>
<td>4.2</td>
<td>3.9</td>
<td></td>
</tr>
</tbody>
</table>
Samples and real images (LSUN churches, 64×64)

Thank you for listening. Come see us at poster 71 :)