HALI: Hierarchical Adversarially Learned Inference

Negar Rostamzadeh

ACM Webinar
January 18, 2018
Hierarchical Adversarially Learned Inference

Mohamed Ishmael Belghazi, Sai Rajeswar, Olivier Mastropietro, Negar Rostamzadeh, Jovana Mitrovic, Aaron Courville

Paper is on Openreview "submitted to ICLR 2018"
Outline

1. Autoencoder and Reconstruction
2. Variational Inference and Variational Autoencoder
3. GAN: Generative Adversarial Networks
4. ALI: Adversarially Learned Inference
5. HALI: Hierarchical Adversarially Learned Inference
6. Results
7. Questions/Answers?!
Autoencoder and Reconstruction
Autoencoder and Reconstruction

Autoencoder

\[ p(z|x) = \frac{p(x|z)p(z)}{p(x)} = \frac{p(x,z)}{p(x)} \]

\[ p(x) = \int p(x|z)p(z)dz \quad \text{likely to be interactable} \]
Variational Inference and Variational Autoencoder
Variational Inference and Variational Autoencoder

\[
\log(p(x, z)) = \log(p(z | x)) + \log(p(x)) \\
\log(p(x)) = \log(p(x, z)) - \log(p(z | x)) \\
\log(p(x)) = \log\left(\frac{p(x, z)}{q(z | x)}\right) + \log\left(\frac{q(z | x)}{p(z | x)}\right) \\
\log(p(x)) = \mathbb{E}_{z \sim q(z | x)}[\log\left(\frac{p(x, z)}{q(z | x)}\right)] + KL(q(z | x) \parallel p(z | x)) \\
\log(p(x)) \geq \mathbb{E}_{z \sim q(z | x)}[\log\left(\frac{p(x, z)}{q(z | x)}\right)] \\
\log(p(x)) \geq \mathbb{E}_{z \sim q(z | x)}[\log\left(\frac{p(x | z)p(z)}{q(z | x)}\right)] \\
\log(p(x)) \geq \mathbb{E}_{z \sim q(z | x)}[\log(p(x | z))] - KL(q(z | x) \parallel p(z))
\]
GAN: Generative Adversarial Networks
GAN: Generative Adversarial Networks\textsuperscript{2}

\begin{figure}
\centering
\includegraphics[width=\textwidth]{fig1.png}
\caption{GAN\textsuperscript{1}}
\end{figure}

\textsuperscript{1}Graphs are taken from Ishmael Belghazi’s blog post/ALI paper with his permission
\textsuperscript{2}GAN: "Generative Adversarial Nets.", Goodfellow et al, NIPS, 2014.
GAN: Generative Adversarial Networks

\[
\begin{align*}
\min_G \max_D V(D, G) &= \mathbb{E}_{q(x)}[\log(D(x))] + \mathbb{E}_{p(z)}[\log(1 - D(G(z)))] \\
&= \int q(x) \log(D(x)) dx \\
&\quad + \int \int p(z)p(x \mid z) \log(1 - D(x)) dx dz. 
\end{align*}
\]
ALI: Adversarially Learned Inference
ALI: Adversarially Learned Inference$^3$, $^4$

- It is a **Deep Directed Generative Model**
- It jointly learns a **Generative** network and an **Inference** network using an adversarial process.
- Unlike the VAE, the objective function involves **no explicit reconstruction loop**.
- ALI tends to produce **believable reconstructions with interesting variations**, instead of **pixel-perfect reconstruction**

---

$^3$ALI: Adversarially Learned Inference, Vincent Dumoulin, Ishmael Belghazi, Ben Poole, Olivier Mastropietro, Alex Lamb, Martin Arjovsky, Aaron Courville

$^4$Adversarial Feature Learning, Jeff Donahue, Philipp Krähenbühl, Trevor Darrell
ALI: Adversarially Learned Inference

\[ z \sim q(z \mid x) \]

\[ G_z(x) \]

\[ x \sim q(x) \]

\[ D(x, z) \]

\[ x, z \sim q(x, z) \]?

\[ 1 / 0 \]

\[ z \sim p(z) \]

\[ G(z) \]

\[ x \sim p(x \mid z) \]
ALI: Adversarially Learned Inference

Consider the two following probability distributions over $x$ and $z$:
- the *encoder* joint distribution $q(x, z) = q(x)q(z \mid x)$,
- the *decoder* joint distribution $p(x, z) = p(z)p(x \mid z)$.

$$\min_G \max_D V(D, G) = \mathbb{E}_{q(x)}[\log(D(x, G_x(z))))] + \mathbb{E}_{p(z)}[\log(1 - D(G_x(z), z)))]$$

$$= \int \int q(x)q(z \mid x) \log(D(x, z))dxdz$$

$$+ \int \int p(z)p(x \mid z) \log(1 - D(x, z))dxdz.$$  (2)
(a) Tiny ImageNet samples.  
(b) Tiny ImageNet reconstructions.

Figure: Samples and reconstructions on the Tiny ImageNet dataset. For the reconstructions, odd columns are original samples from the validation set and even columns are corresponding reconstructions.
ALI- SVHN: Samples and Reconstruction

(a) SVHN samples.  
(b) SVHN reconstructions.

Figure: Samples and reconstructions on the SVHN dataset. For the reconstructions, odd columns are original samples from the validation set and even columns are corresponding reconstructions.
ALI- CIFAR10: Samples and Reconstruction

(a) CIFAR10 samples.  
(b) CIFAR10 reconstructions.

Figure: Samples and reconstructions on the CIFAR10 dataset. For the reconstructions, odd columns are original samples from the validation set and even columns are corresponding reconstructions.
ALI- CelebA: Samples and Reconstruction

(a) CelebA samples.  
(b) CelebA reconstructions.

Figure: Samples and reconstructions on the CelebA dataset. For the reconstructions, odd columns are original samples from the validation set and even columns are corresponding reconstructions.
ALI - Latent space interpolation

Figure: Latent space interpolations on the CelebA validation set. Left and right columns correspond to the original pairs $x_1$ and $x_2$, and the columns in between correspond to the decoding of latent representations interpolated linearly from $z_1$ to $z_2$. Unlike other adversarial approaches like DCGAN, ALI allows one to interpolate between actual data points.
### Table: SVHN test set misclassification rate

<table>
<thead>
<tr>
<th>Model</th>
<th>Misclassification rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>VAE (M1 + M2)</td>
<td>36.02</td>
</tr>
<tr>
<td>SWWAE with dropout</td>
<td>23.56</td>
</tr>
<tr>
<td>DCGAN + L2-SVM</td>
<td>22.18</td>
</tr>
<tr>
<td>SDGM</td>
<td>16.61</td>
</tr>
<tr>
<td><strong>GAN (feature matching)</strong></td>
<td><strong>8.11 ± 1.3</strong></td>
</tr>
<tr>
<td>ALI (ours, L2-SVM)</td>
<td>19.14 ± 0.50</td>
</tr>
<tr>
<td><strong>ALI (ours, no feature matching)</strong></td>
<td><strong>7.42 ± 0.65</strong></td>
</tr>
</tbody>
</table>
Table: CIFAR10 test set missclassification rate for semi-supervised learning using different numbers of trained labeled examples. For ALI, error bars correspond to 3 times the standard deviation.

<table>
<thead>
<tr>
<th>Model</th>
<th>Number of labeled examples</th>
<th>1000</th>
<th>2000</th>
<th>4000</th>
<th>8000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ladder network</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CatGAN</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GAN (feature matching)</td>
<td>21.83 ± 2.01</td>
<td>19.61 ± 2.09</td>
<td>18.63 ± 2.32</td>
<td>17.72 ± 1.82</td>
<td></td>
</tr>
<tr>
<td>ALI (ours, no feature matching)</td>
<td>19.98 ± 0.89</td>
<td>19.09 ± 0.44</td>
<td>17.99 ± 1.62</td>
<td>17.05 ± 1.49</td>
<td></td>
</tr>
</tbody>
</table>
**ALI- Conditional Generation**

<table>
<thead>
<tr>
<th>(a)</th>
<th>(b)</th>
<th>(c)</th>
<th>(d)</th>
<th>(e)</th>
<th>(f)</th>
<th>(g)</th>
<th>(h)</th>
<th>(i)</th>
<th>(j)</th>
<th>(k)</th>
<th>(l)</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>II</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>III</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IV</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Figure:** The attributes are male, attractive, young for row I; male, attractive, older for row II; female, attractive, young for row III; female, attractive, older for Row IV. Attributes are then varied uniformly over rows across all columns in the following sequence: (b) black hair; (c) brown hair; (d) blond hair; (e) black hair, wavy hair; (f) blond hair, bangs; (g) blond hair, receding hairline; (h) blond hair, balding; (i) black hair, smiling; (j) black hair, smiling, mouth slightly open; (k) black hair, smiling, mouth slightly open, eyeglasses; (l) black hair, smiling, mouth slightly open, eyeglasses, wearing hat.
HALI: Hierarchical Adversarially Learned Inference
What is HALI?

- HALI is a hierarchical Generative model with a Markovian structure.
- It jointly trains generative and inference model.

HALI provides ...

- semantically meaningful reconstructions with different levels of fidelity.
- progressively more abstract latent representations.
- useful representation for downstream tasks.
HALI: Hierarchical Adversarially Learned Inference

The encoder and decoder distributions:

Joint distribution of the encoder:

$$q(x, \ldots, z_L) = \prod_{l=2}^{L} q(z_l | z_{l-1}) q(z_1 | x) q(x),$$  \hspace{1cm} (3)

Joint distribution of the decoder:

$$p(x, \ldots, z_L) = p(x | z_1) \prod_{l=2}^{L} p(z_{l-1} | z_l) p(z_L).$$ \hspace{1cm} (4)
HALI: Hierarchical Adversarially Learned Inference

\[
\mathcal{L}^l(x) = \mathbb{E}_{z_l \sim T_{z_l} | x} \left[ - \log(p(x | z_l)) \right]
\]
HALI: Hierarchical Adversarially Learned Inference
HALI vs ALI

- Both relies on joint training of the generative and inference models.
- HALI leverages the hierarchical architecture to:
  - Offer reconstruction of the same data sample with increasing levels of fidelity.
  - Abstraction of learned representation increases as we go up the hierarchy.
  - Flexible inference model that provides useful representations for downstream tasks.
Results
Qualitative Results - SVHN - Reconstruction

(a) SVHN from $z_1$

(b) SVHN from $z_2$

Figure: Reconstructions for SVHN from $z_1$ and reconstructions from $z_2$. Odd columns corresponds to examples from the validation set while even columns are the model’s reconstructions.
Qualitative Results - CIFAR10 - Reconstruction

(a) CIFAR10 from $z_1$

(b) CIFAR10 from $z_2$

Figure: Reconstructions for CIFAR10 from $z_1$ and reconstructions from $z_2$. Odd columns corresponds to examples from the validation set while even columns are the model’s reconstructions.
Qualitative Results - Imagenet128 - Reconstruction

(a) ImageNet128 from $z_1$

(b) ImageNet128 from $z_2$

Figure: ImageNet128 reconstructions from $z_1$ and $z_2$. Odd columns corresponds to examples from the validation set while even columns are the model’s reconstructions.
Qualitative Results - Imagenet128 - Samples

(a) ImageNet128

Figure: Samples from 128 × 128 ImageNet128 dataset
Qualitative Results - CelebA - Samples

Figure: Samples from $128 \times 128$ CelebA dataset

(a) CelebA
Quality of the reconstruction: HALI vs ALI

Table: Summary of CelebA attributes classification from reconstructions for VAE, ALI and the two levels of HALI.

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Std</th>
<th># Best</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data</td>
<td>77.13</td>
<td>12.48</td>
<td></td>
</tr>
<tr>
<td>VAE</td>
<td>81.28</td>
<td>10.50</td>
<td>5</td>
</tr>
<tr>
<td>ALI</td>
<td>84.60</td>
<td>5.73</td>
<td>3</td>
</tr>
<tr>
<td>HALI $z_1$</td>
<td>91.35</td>
<td>5.62</td>
<td>27</td>
</tr>
<tr>
<td>HALI $z_2$</td>
<td>86.28</td>
<td>5.64</td>
<td>3</td>
</tr>
</tbody>
</table>
Perceptual Reconstructions

Figure: Comparison of average reconstruction error over the validation set for each level of reconstructions using the Euclidean (a) and discriminator embedded (b) distances.

---

Figure: Inpainting on center cropped images on CelebA
<table>
<thead>
<tr>
<th>Method</th>
<th>MNIST (# errors)</th>
</tr>
</thead>
<tbody>
<tr>
<td>VAE (M1+M2) [kingma et al, 2014]</td>
<td>233 ± 14</td>
</tr>
<tr>
<td>VAT [Miyato et al, 2017]</td>
<td>136</td>
</tr>
<tr>
<td>CatGAN</td>
<td>191 ± 10</td>
</tr>
<tr>
<td>Adversarial Autoencoder [makhzani et al, 2015]</td>
<td>190 ± 10</td>
</tr>
<tr>
<td>PixelGAN [makhzani et al, 2017]</td>
<td>108 ± 15</td>
</tr>
<tr>
<td>ADGM [Maaloe et al, 2016]</td>
<td>96 ± 2</td>
</tr>
<tr>
<td>Feature-Matching GAN [Salimans et al, 2016]</td>
<td>93 ± 6.5</td>
</tr>
<tr>
<td>GSSLTRABG [dai et al, 2017]</td>
<td>79.5 ± 9.8</td>
</tr>
<tr>
<td>HALI (ours)</td>
<td><strong>73</strong></td>
</tr>
</tbody>
</table>

**Table:** Comparison on semi-supervised learning with state-of-the-art methods on MNIST with 100 labels instance per class. Only methods without data augmentation are included.
Figure: Inpainting on center cropped images on SVHN
**Figure:** Inpainting on center cropped images on MS-COCO dataset
Figure: Real CelebA faces (right) and their corresponding innovation tensor (IT) updates (left). For instance, the third row in the figure features Christina Hendricks followed by hair-color IT updates. Similarly, the first two rows depicts usage of smile-IT and the 4th row glasses-plus-hair-color-IT.
Questions/Answers?!
Thanks!