Generative Model Variants

-- GLO, IMLE, GLANN

Hao Dong

Peking University

Generative Model Variants -- GLO, IMLE, GLANN

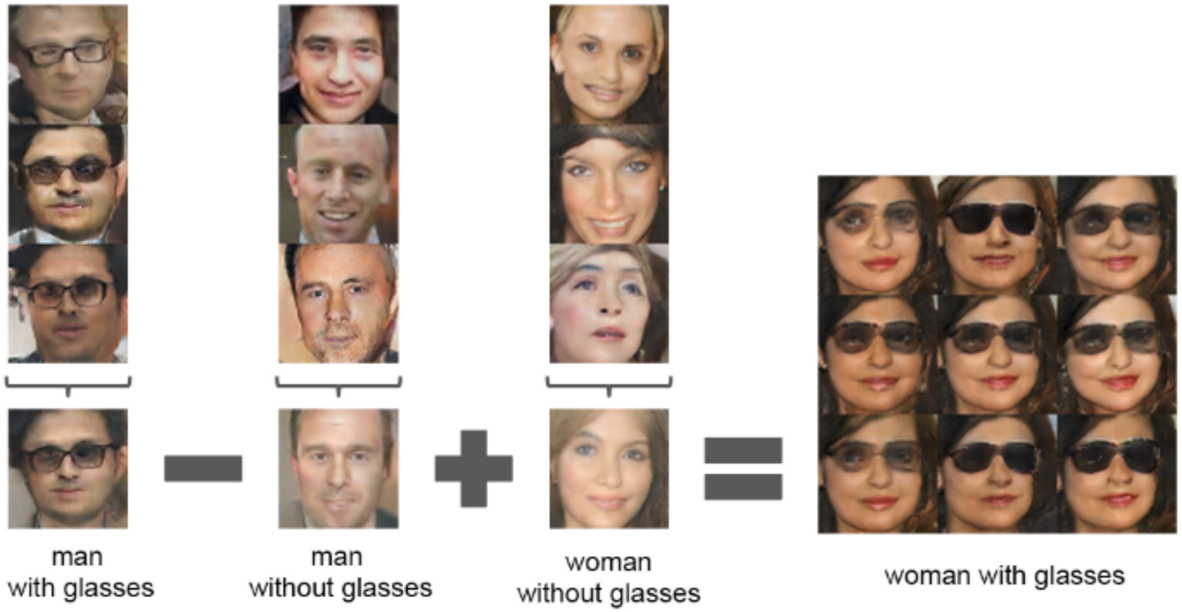
北京大學
PEKING UNIVERSITY

- Generative Latent Optimisation, GLO
- Implicit Maximum Likelihood Estimation, IMLE
- Generative Latent Nearest Neighbors, GLANN
- Discussion

Recap: DCGAN

- **Properties of DCGAN**

  - Generation from noise: synthesis new images from data distribution
  - Interpolation: translate linear interpolation in noise space to semantic interpolation in image
  - Linear operation: linear arithmetic in noise space



Bojanowski, P., Joulin, A., Lopez, D., & Arthur, P. (2018). GLO: Optimizing the Latent Space of Generative Networks.

Recap: DCGAN

- **Properties of DCGAN**
  - Mode collapse: generator may model only a few localised regions of data distribution
  - Adversarial training: unstable for training, sensitive to initialisation, architecture and hyper parameters

- **GLO**
- **IMLE**
- **GLANN**

Aim at solving major problems in GANs to develop NEW generative models.

Bojanowski, P., Joulin, A., Lopez, D., & Arthur, P. (2018). GLO: Optimizing the Latent Space of Generative Networks.

- Generative Latent Optimisation, GLO
- Implicit Maximum Likelihood Estimation, IMLE
- Generative Latent Nearest Neighbors, GLANN
- Discussion

Generative Latent Optimisation (GLO)

"**Success of GANs** comes from:
- Inductive bias of deep convolutional networks
- The adversarial training protocol "

GLO: aims at preserving good properties of GAN, without adversarial training and mode collapse

Approach:
- Image space $\mathcal{X}$, noise space $\mathcal{Z} \subset \mathbb{R}^d$. Training set of images $\{x_1, \ldots, x_N\}$.
- For each image x_i, take a random noise $z_i \in \mathcal{Z}$.
- Construct all the pairs: $\{(z_i, x_i)\}_{i=1}^N$
- Target is to learn the mapping $g_\theta: \mathcal{Z} \to \mathcal{X}$, where each training image x_i is mapped from its noise z_i

Bojanowski, P., Joulin, A., Lopez, D., & Arthur, P. (2018). GLO: Optimizing the Latent Space of Generative Networks.

Generative Latent Optimisation (GLO)

PEKING UNIVERSITY 北京大學

Approach:

- Parameterize a generator $g_\theta : \mathcal{Z} \to \mathcal{X}$ using CNN
- Initialize θ, $\{z_i\}$
- Training objective:

$$\min_\theta \sum_{i=1}^{N} [\min_{\mathrm{z_i} \in \mathcal{Z}} \ell(g_\theta(z_i), x_i)]$$

Notice:

- Noise z_i are learnable, jointly optimised with θ, to form the structure of the noise space.
- It is like a "non-parameterized autoencoder".

Bojanowski, P., Joulin, A., Lopez, D., & Arthur, P. (2018). GLO: Optimizing the Latent Space of Generative Networks.

Generative Latent Optimisation (GLO)

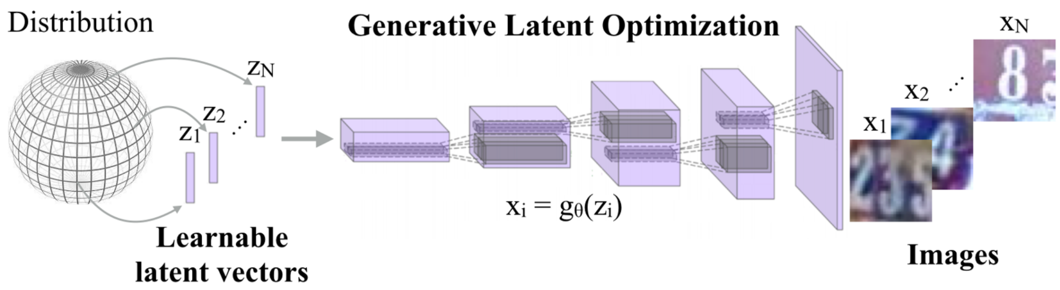
Noise space

- A common choice of $\mathcal{Z}$ in the GANs literature is unit Gaussian distribution from $\mathbb{R}^d$

$$z \sim \mathcal{N}(0, I)$$

- z drawn from unit Gaussian distribution is unlikely to land far outside the sphere of radius $\sqrt{d}$, i.e. with high probability, $\|z\|_2 \in (\sqrt{d} - \varepsilon, \sqrt{d} + \varepsilon)$.

- Instead of using Gaussian, GLO use the unit sphere as the noise space.



Bojanowski, P., Joulin, A., Lopez, D., & Arthur, P. (2018). GLO: Optimizing the Latent Space of Generative Networks.

Generative Latent Optimisation (GLO)

Noise space

Think: why not using a Gaussian distribution?

Because in the optimisation objective, the noise vectors need to be optimised. It is easy to constrain them into a sphere, by optimising and projecting them back to the sphere for each step.

Bojanowski, P., Joulin, A., Lopez, D., & Arthur, P. (2018). GLO: Optimizing the Latent Space of Generative Networks.

Generative Latent Optimisation (GLO)

Loss function

- Simple choice: squared error

$$\ell_2(x, x') = \|x - x'\|_2^2$$

- Leads to blurry reconstructions of natural images (moving to the average)

- GANs generate sharp image, because they use a CNN (discriminator) as the loss function, whose early layers focus on edges.

- GLO uses Laplacian pyramid loss for sharp reconstruction.

Bojanowski, P., Joulin, A., Lopez, D., & Arthur, P. (2018). GLO: Optimizing the Latent Space of Generative Networks.

Generative Latent Optimisation (GLO)

Loss function

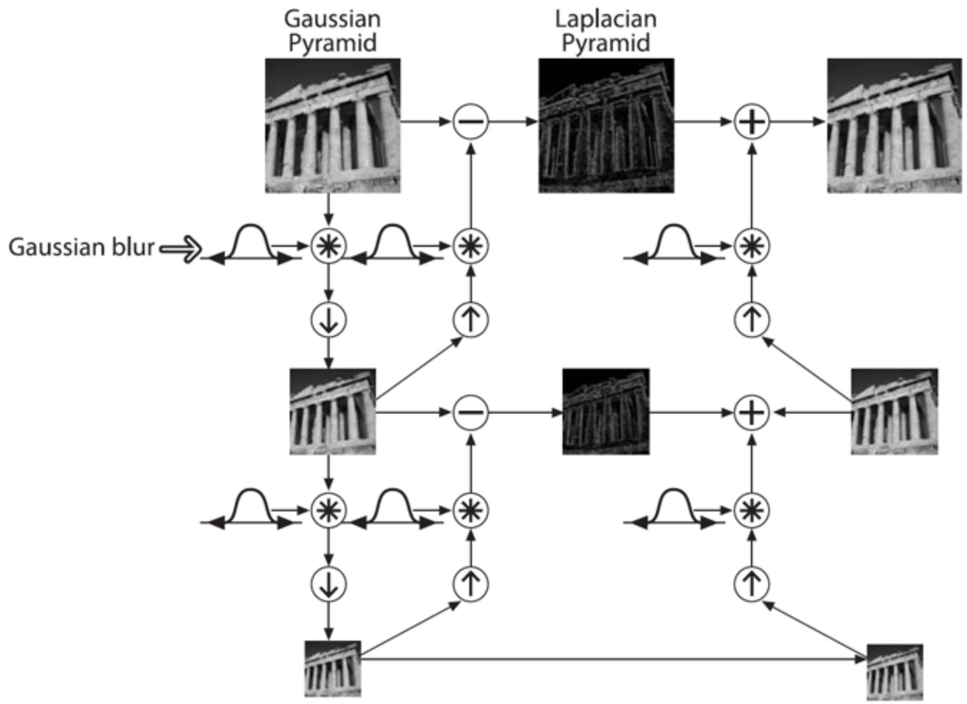
- Laplacian pyramid loss:

$$Lap(x, x') = \sum_j 2^{2j} \left| L^j(x) - L^j(x') \right|_1$$

- L^j is the j^{th} level residual in the image's Laplacian pyramid, which is the residual between the j^{th} level image and the upsampled $j+1^{th}$ level, reflecting the details lost during upsampling.

$$L^j(x) = G^j(x) - Upsample(G^{j+1}(x))$$

G^j is the j^{th} downsampled image in the Gaussian pyramid.



Bojanowski, P., Joulin, A., Lopez, D., & Arthur, P. (2018). GLO: Optimizing the Latent Space of Generative Networks.

Generative Latent Optimisation (GLO)

Discussion

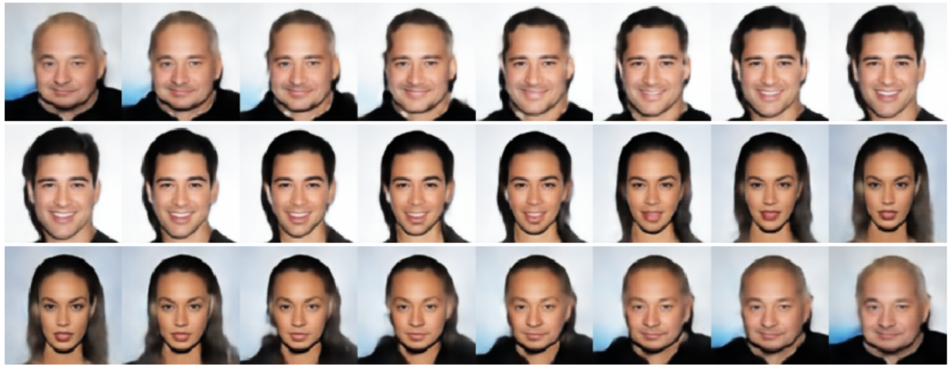
- Generation
  - Distribution of noise on the unit sphere is unknown
  - Need to fit another (simple) model to map a known distribution to the noise distribution
  - E.g. fit a full-covariance Gaussian distribution

- Properties of latent interpolation & linear operation are preserved.

- Mode Collapse
  - Issue of mode collapse is addressed, because all training samples are embedded in the noise space.

Bojanowski, P., Joulin, A., Lopez, D., & Arthur, P. (2018). GLO: Optimizing the Latent Space of Generative Networks.

Generative Latent Optimisation (GLO)



- Nice result on interpolation for images in the training set.
- GLO cannot outperform DCGAN on image generation.
    - Noise space for GLO is more like a feature embedding space, not for sampling.

(d) CelebA-128

(f) LSUN-128

Bojanowski, P., Joulin, A., Lopez, D., & Arthur, P. (2018). GLO: Optimizing the Latent Space of Generative Networks.

- Generative Latent Optimisation, GLO
- Implicit Maximum Likelihood Estimation, IMLE
- Generative Latent Nearest Neighbors, GLANN
- Discussion

Implicit Maximum Likelihood Estimation (IMLE)

Recall

- Generative models aim at model the data distribution $x \sim p_{data}$
- A common approach is to
  - sample from a known distribution
    $$z \sim q(z)$$
  - then use a parametric function to map the known distribution to the data distribution
    $$x = G_{\theta}(z)$$
- Denote the learned distribution $x \sim p_{\theta}$

Li, K., Malik, J., Sciences, C., & States, U. (2018). IMLE: Implicit Maximum Likelihood Estimation.

Implicit Maximum Likelihood Estimation (IMLE)

Recall

- Maximum likelihood estimation (MLE)
  - Maximise the probability of p_θ to generate the data sampled from p_{data}

- MLE is equivalent to minimising KL divergence between two distributions

$$\min_\theta D_{KL}(p_{data}||p_\theta) = \min_\theta \int_x p_{data} \log(\frac{p_{data}}{p_\theta})\, dx$$

- However, the probability density function of p_θ is always intractable

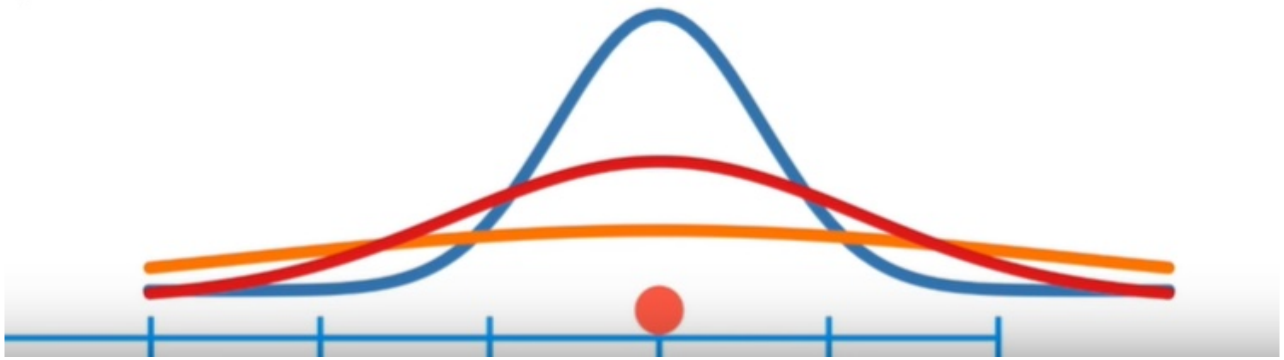
$$p_\theta(x) = \frac{\partial}{\partial x_1} \dots \frac{\partial}{\partial x_d} \int_{\{z|\forall i, \left(G_\theta(z)\right)_i \leq x_i\}} q(z)\, dz$$

Li, K., Malik, J., Sciences, C., & States, U. (2018). IMLE: Implicit Maximum Likelihood Estimation.

Implicit Maximum Likelihood Estimation (IMLE)

Intuition

- MLE should assign high probability density to each of the samples



- We can adjust the parameter θ so that samples drawn from p_θ are close to samples drawn from p_{data}

Li, K., Malik, J., Sciences, C., & States, U. (2018). IMLE: Implicit Maximum Likelihood Estimation.

Implicit Maximum Likelihood Estimation (IMLE)

Algorithm

Require: The dataset $D = \{\mathbf{x}_i\}_{i=1}^n$ and a sampling mechanism for the implicit model P_θ

Initialize θ to a random vector

for $k = 1$ **to** K **do**

  Draw i.i.d. samples $\tilde{\mathbf{x}}_1^\theta, \ldots, \tilde{\mathbf{x}}_m^\theta$ from P_θ

  Pick a random batch $S \subseteq \{1, \ldots, n\}$

  $\sigma(i) \leftarrow \arg\min_j \left\| \mathbf{x}_i - \tilde{\mathbf{x}}_j^\theta \right\|_2^2 \ \forall i \in S$

  for $l = 1$ **to** L **do**

    Pick a random mini-batch $\tilde{S} \subseteq S$

    $\theta \leftarrow \theta - \eta \nabla_\theta \left(\frac{n}{|\tilde{S}|} \sum_{i \in \tilde{S}} \left\| \mathbf{x}_i - \tilde{\mathbf{x}}_{\sigma(i)}^\theta \right\|_2^2 \right)$

  end for

end for

return θ

Li, K., Malik, J., Sciences, C., & States, U. (2018). IMLE: Implicit Maximum Likelihood Estimation.

Implicit Maximum Likelihood Estimation (IMLE)

Algorithm analysis

- Mode collapse is addressed, because the training process encourages the generated samples to be close to all the samples in training data.
- The training objective is a simple l_2 loss. It is easy to optimise.
- Distance measurement to compute the nearest neighbor and to optimise is critical. L_2 distance used in paper leads to generating blur images.

Li, K., Malik, J., Sciences, C., & States, U. (2018). IMLE: Implicit Maximum Likelihood Estimation.

Implicit Maximum Likelihood Estimation (IMLE)

Algorithm analysis

- L_2 distance used in paper leads to blur results.

IMLE trained on cifar-10.

Li, K., Malik, J., Sciences, C., & States, U. (2018). IMLE: Implicit Maximum Likelihood Estimation.

Implicit Maximum Likelihood Estimation (IMLE)

Theoretical analysis

Why does this simple algorithm work?

- IMLE has a theoretical proof, showing the algorithm is equivalent to MLE.

Recap:

- $z \sim q(z)$
- $x = G_\theta(z)$
- $p_\theta(x) = \frac{\partial}{\partial x_1} \dots \frac{\partial}{\partial x_d} \int_{\{z \mid \forall i, \, (G_\theta(z))_i \leq x_i\}} q(z)\, dz$
- MLE Objective: $\min_\theta D_{KL}(p_{data} || p_\theta)$

Li, K., Malik, J., Sciences, C., & States, U. (2018). IMLE: Implicit Maximum Likelihood Estimation.

Implicit Maximum Likelihood Estimation (IMLE)

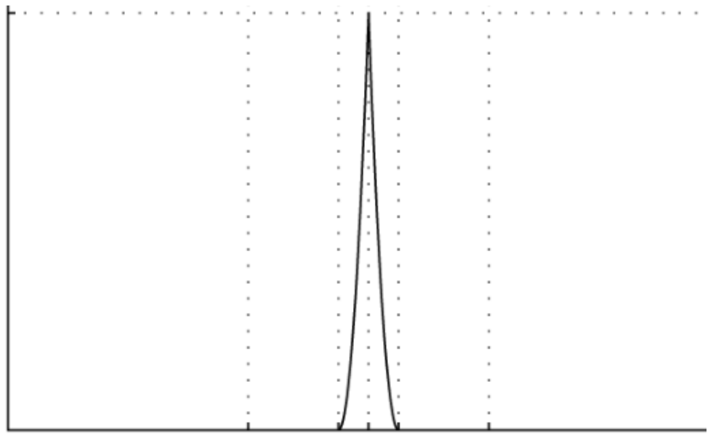
Theoretical Analysis

Lemma : $p_\theta(x) = \int \delta\big(\mathrm{x} - G(z)\big)q(z)dz$, where δ is the Dirac function.

Dirac δ function is a generalised function satisfies:

$$\delta(x) = \begin{cases} 0, x \neq 0 \\ \infty, x = 0 \end{cases}$$

$$\int_{-\infty}^{+\infty} \delta(x) = 1$$



Property: for any function f, $\int_{-\infty}^{+\infty} f(x+y)\delta(x)\,dx = f(y)$

Li, K., Malik, J., Sciences, C., & States, U. (2018). IMLE: Implicit Maximum Likelihood Estimation.

Implicit Maximum Likelihood Estimation (IMLE)

PEKING UNIVERSITY

Theoretical Analysis

Lemma : $p_\theta(x) = \int \delta\big(\mathrm{x} - G(z)\big) q(z) dz$, where δ is the Dirac function.

Proof:

$$
\begin{aligned}
&\int_{-\infty}^{x} dt \int \delta\big(t - G(z)\big) q(z) dz \\
&= \int \left\{ q(z) dz \int_{-\infty}^{x} \delta\big(t - G(z)\big)\, dt \right\} \\
&= \int \left\{ q(z) dz\, \boldsymbol{I}\left[\forall i, \big(G_\theta(z)\big)_i < (x)_i\right]\right\} \\
&= \int_{\{z \mid \forall i, \left(G_\theta(z)\right)_i \le x_i\}} q(z) dz \\
&= \int_{-\infty}^{x} p_\theta(x) dx
\end{aligned}
$$

Li, K., Malik, J., Sciences, C., & States, U. (2018). IMLE: Implicit Maximum Likelihood Estimation.

Implicit Maximum Likelihood Estimation (IMLE)

Theoretical Analysis

$$p_\theta(x) = \int \delta\big(\mathrm{x} - G(z)\big)q(z)dz = \mathbb{E}_{z\sim q(z)}[\delta(x - G(z))]$$

Dirac function is also a limit of isotropic Gaussian distribution:

$$\delta(x) = \lim_{\sigma\to 0}\frac{1}{(2\pi\sigma^2)^{d/2}}\exp(-\frac{\|x\|^2}{2\sigma^2})$$

Then,

$$p_\theta(x) = \lim_{\sigma\to 0}\mathbb{E}_{z\sim q(z)}\left[\frac{1}{(2\pi\sigma^2)^{\frac{d}{2}}}\exp\left(-\frac{\|x - G(z)\|^2}{2\sigma^2}\right)\right]$$

Li, K., Malik, J., Sciences, C., & States, U. (2018). IMLE: Implicit Maximum Likelihood Estimation.

Implicit Maximum Likelihood Estimation (IMLE)

Theoretical Analysis

MLE Objective:

$$\underset{\theta}{\text{minimise}}\, D_{KL}(p_{data}||p_{\theta})$$

$$\Leftrightarrow \underset{\theta}{\text{minimise}} \int_x p_{data} \log(\frac{p_{data}}{p_{\theta}})\, dx$$

$$\Leftrightarrow \underset{\theta}{\text{minimise}}\, \mathbb{E}_{x\sim p_{data}(x)}[-\log(p_{\theta})]$$

$$\Leftrightarrow \underset{\theta}{\text{minimise}}\, \mathbb{E}_{x\sim p_{data}(x)}\left[-\log\left\{\lim_{\sigma\to 0} \mathbb{E}_{z\sim q(z)}\left[\frac{1}{(2\pi\sigma^2)^{\frac{d}{2}}}\exp\left(-\frac{\|x-G(z)\|^2}{2\sigma^2}\right)\right]\right\}\right]$$

$$\Leftrightarrow \underset{\theta}{\text{minimise}} \lim_{\sigma\to 0} \mathbb{E}_{x\sim p_{data}(x)}\left[-\log\left\{\mathbb{E}_{z\sim q(z)}\left[\exp\left(-\frac{\|x-G(z)\|^2}{2\sigma^2}\right)\right]\right\}\right]$$

Li, K., Malik, J., Sciences, C., & States, U. (2018). IMLE: Implicit Maximum Likelihood Estimation.

Implicit Maximum Likelihood Estimation (IMLE)

Theoretical Analysis

MLE Objective

- In practice, we sample the data from distribution to approximate the expectation
- Suppose $x_1, x_2, \ldots, x_M$ are drawn from p_{data} (training data); $z_1, z_2, \ldots, z_N$ are drawn from $q(z)$

$$\underset{\theta}{\text{minimise}} \lim_{\sigma \to 0} \mathbb{E}_{x \sim p_{data}(x)} \left[-\log \left\{ \mathbb{E}_{z \sim q(z)} \left[\exp \left(-\frac{\|x - G(z)\|^2}{2\sigma^2} \right) \right] \right\} \right]$$

$$\Leftrightarrow \underset{\theta}{\text{minimise}} \lim_{\sigma \to 0} -\frac{1}{M} \sum_{i=1}^{\mathrm{M}} \log \left\{ \sum_{j=1}^{N} exp \left(-\frac{\left\| x_i - G\left(z_j\right) \right\|^2}{2\sigma^2} \right) \right\}$$

Li, K., Malik, J., Sciences, C., & States, U. (2018). IMLE: Implicit Maximum Likelihood Estimation.

Implicit Maximum Likelihood Estimation (IMLE)

Theoretical Analysis

MLE Objective

$$\underset{\theta}{\text{minimise}} \lim_{\sigma \to 0} -\frac{1}{M} \sum_{i=1}^{M} \log \left\{ \sum_{j=1}^{N} exp\left(-\frac{\left\| x_i - G(z_j) \right\|^2}{2\sigma^2} \right) \right\}$$

$$\Leftrightarrow \underset{\theta}{\text{minimise}} - \sum_{i=1}^{M} max_{j=1}^{N} \left\{ -\left\| x_i - G(z_j) \right\|^2 \right\}$$

$$\Leftrightarrow \underset{\theta}{\text{minimise}} \sum_{i=1}^{M} min_{j=1}^{N} \left\{ \left\| x_i - G(z_j) \right\|^2 \right\}$$

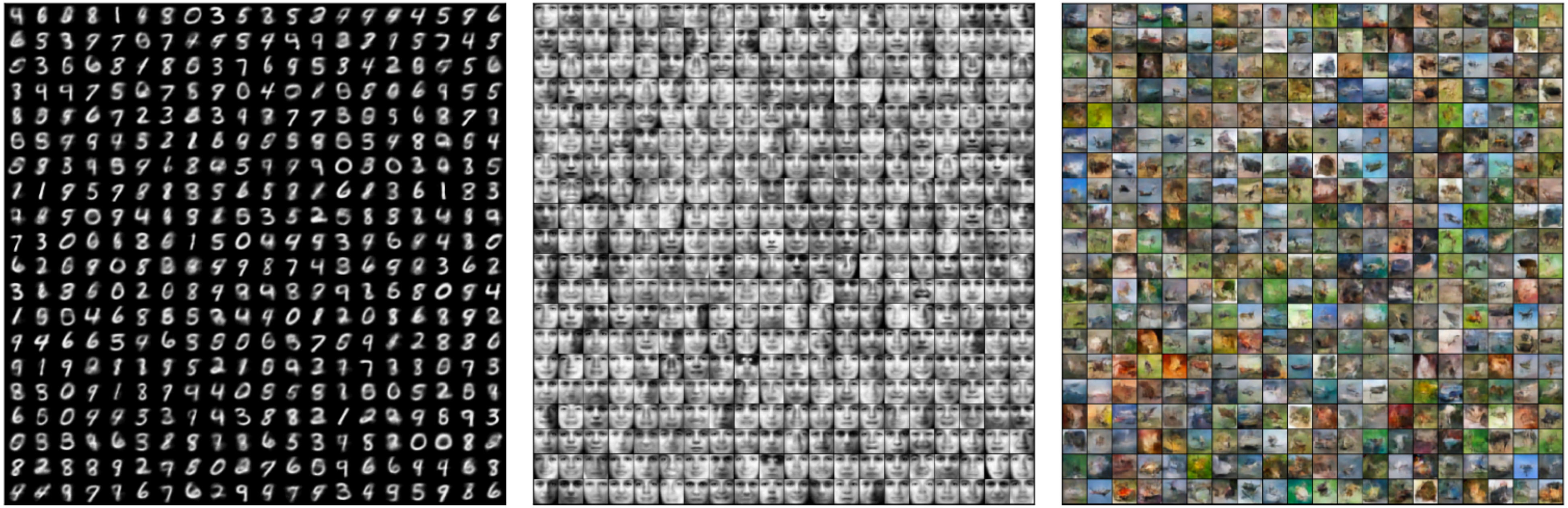
Minimise the nearest neighbor distance

Li, K., Malik, J., Sciences, C., & States, U. (2018). IMLE: Implicit Maximum Likelihood Estimation.

Implicit Maximum Likelihood Estimation (IMLE)

Results

- L_2 distance leads to **blur images.**



(a) MNIST (b) TFD (c) CIFAR-10

Li, K., Malik, J., Sciences, C., & States, U. (2018). IMLE: Implicit Maximum Likelihood Estimation.

- Generative Latent Optimisation, GLO
- Implicit Maximum Likelihood Estimation, IMLE
- Generative Latent Nearest Neighbors, GLANN
- Discussion

Generative Latent Nearest Neighbors (GLANN)

Main Drawbacks of GLO and IMLE

- GLO
  - The distribution of noise space Z is unknown.
  - There is no principled way to sample new images.
  - Fitting Gaussian to the noise distribution in the paper does not synthesis high quality images.

- IMLE
  - IMLE is sensitive to the distance metric used.
  - Using L_2 distance on image pixel space causes blurry synthesised images.
  - Computing nearest neighbour for each sample in the image space is costly.

Hoshen, Y., Li, K., Malik, J., & Berkeley, U. C. (2019). GLANN: Non-Adversarial Image Synthesis with Generative Latent Nearest Neighbors.

Generative Latent Nearest Neighbors (GLANN)

Approach

- Linear arithmetic in GLO’s noise space can semantically manipulate the image.
- That means Euclidean metric in GLO’s noise space is semantically meaningful.
- Why not use GLO to learn a good latent distribution, then apply IMLE to sample from noise?

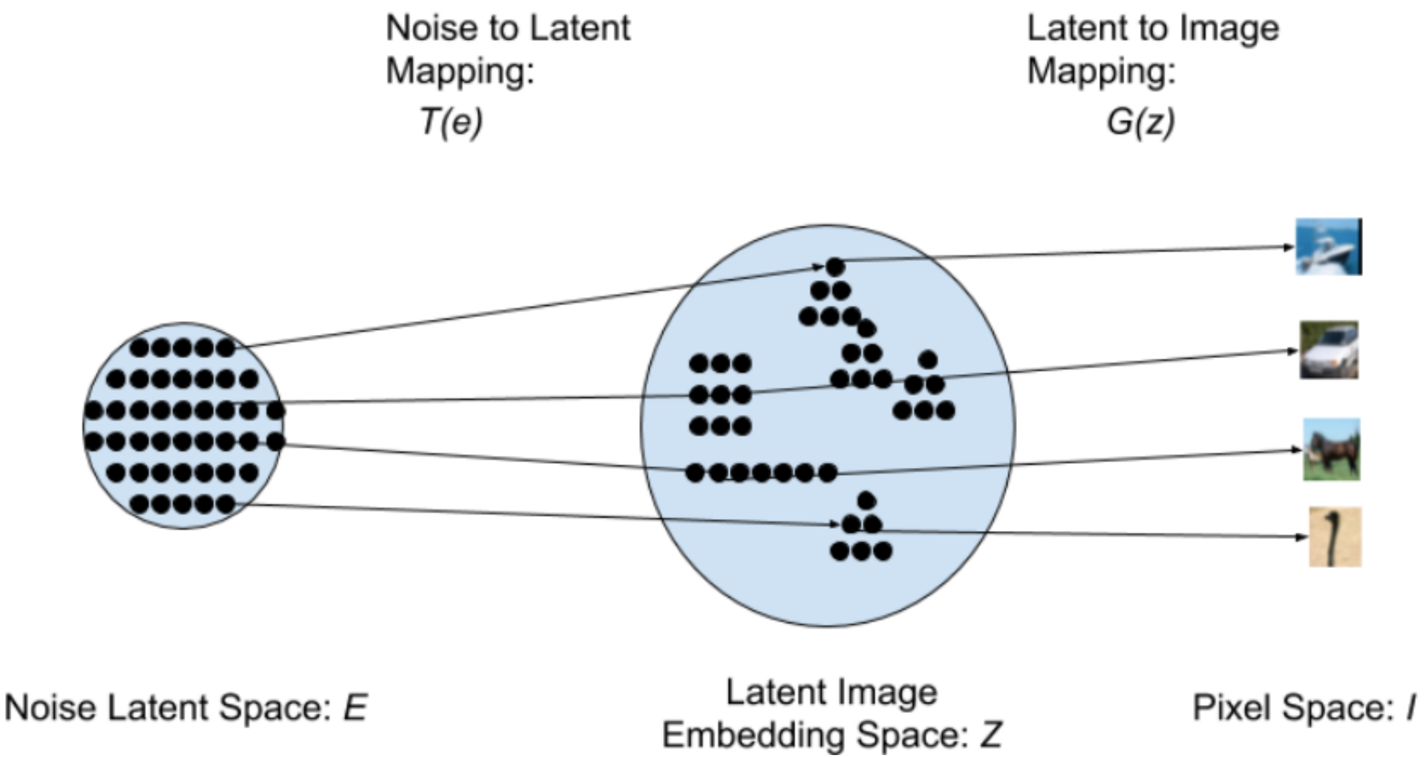
Hoshen, Y., Li, K., Malik, J., & Berkeley, U. C. (2019). GLANN: Non-Adversarial Image Synthesis with Generative Latent Nearest Neighbors.

Generative Latent Nearest Neighbors (GLANN)

Approach

Stage 2: IMLE learns e to z

Stage 1: GLO learns z to X



Hoshen, Y., Li, K., Malik, J., & Berkeley, U. C. (2019). GLANN: Non-Adversarial Image Synthesis with Generative Latent Nearest Neighbors.

Generative Latent Nearest Neighbors (GLANN)

Approach

- Stage 1: Latent2Image using GLO
  - $z \mapsto x$
  - Use VGG perceptual loss function

$$\arg\min_{\tilde{G},\{z_i\}} \sum_i \ell_{\text{perceptual}}\left(\tilde{G}(z_i), x_i\right) \quad \text{s.t.} \quad \|z_i\| = 1$$

- Stage 2: Noise2Latent using IMLE
  - $e \longmapsto z$
  - Using Euclidean distance in the latent space

- Sampling new images
  - $e \sim N(0, I)$
  - $\mathrm{e} \longmapsto z \longmapsto \mathrm{x}$

Hoshen, Y., Li, K., Malik, J., & Berkeley, U. C. (2019). GLANN: Non-Adversarial Image Synthesis with Generative Latent Nearest Neighbors.

Generative Latent Nearest Neighbors (GLANN)

PEKING UNIVERSITY

Approach

- Stage 1: Latent2Image using GLO
  - $z \mapsto x$
  - Use VGG perceptual loss function

$$\arg \min_{\tilde{G},\{z_i\}} \sum_i \ell_{\text{perceptual}}\left(\tilde{G}(z_i), x_i\right) \quad \text{s.t.} \quad \|z_i\| = 1$$

$$\tilde{z}_m = T(e_m)$$

- Stage 2: Noise2Latent using IMLE
  - $e \longmapsto z$
  - Using Euclidean distance in the latent space

$$e_t = \arg \min_{e_m} \|z_t - T(e_m)\|_2^2$$

$$T = \arg \min_{\tilde{T}} \sum_t \left\| z_t - \tilde{T}(e_t) \right\|_2^2$$

- Sampling new images
  - $e \sim N(0, I)$
  - $e \longmapsto z \longmapsto x$

Hoshen, Y., Li, K., Malik, J., & Berkeley, U. C. (2019). GLANN: Non-Adversarial Image Synthesis with Generative Latent Nearest Neighbors.

Generative Latent Nearest Neighbors (GLANN)

PEKING UNIVERSITY

Approach

- Stage 1: Latent2Image using GLO
  - $z \mapsto x$
  - Use VGG perceptual loss function

$$\arg\min_{\tilde{G},\{z_i\}} \sum_i \ell_{\text{perceptual}}\left(\tilde{G}\left(z_i\right), x_i\right) \quad \text{s.t.} \quad \|z_i\| = 1$$

$$\tilde{z}_m = T\left(e_m\right)$$

- Stage 2: Nosie2Latent using IMLE
  - $e \mapsto z$
  - Using Euclidean distance in the latent space

$$e_t = \arg\min_{e_m} \|z_t - T\left(e_m\right)\|_2^2$$

$$T = \arg\min_{\tilde{T}} \sum_t \left\|z_t - \tilde{T}\left(e_t\right)\right\|_2^2$$

- Sampling new images
  - $e \sim N(0, I)$
  - $\mathrm{e} \mapsto z \mapsto \mathrm{x}$

Hoshen, Y., Li, K., Malik, J., & Berkeley, U. C. (2019). GLANN: Non-Adversarial Image Synthesis with Generative Latent Nearest Neighbors.

Generative Latent Nearest Neighbors (GLANN)

Results

IMLE GLO GAN GLANN

Hoshen, Y., Li, K., Malik, J., & Berkeley, U. C. (2019). GLANN: Non-Adversarial Image Synthesis with Generative Latent Nearest Neighbors.

Generative Latent Nearest Neighbors (GLANN)

Results

Interpolation on CelebA-HQ 1024*1024

Generating 3D data

Hoshen, Y., Li, K., Malik, J., & Berkeley, U. C. (2019). GLANN: Non-Adversarial Image Synthesis with Generative Latent Nearest Neighbors.

- Generative Latent Optimisation, GLO
- Implicit Maximum Likelihood Estimation, IMLE
- Generative Latent Nearest Neighbors, GLANN
- Discussion

Discussion: What is the Ideal Generative Model?

- Explicit inverse/encode $E: x \to z, G: z \to x$
- Interpolation in latent space (prior distribution)
- Avoid mode collapse
- Fast training
- Generate high-dimensional, high-quality data
- Disentanglement
- …

Summary

- Generative Latent Optimisation, GLO 2018
- Implicit Maximum Likelihood Estimation, IMLE 2018
- Generative Latent Nearest Neighbors, GLANN 2019
- Discussion

北京大學
PEKING UNIVERSITY

Thanks