Glow: Generative Flow with Invertible 1x1 Convolutions

Ким Алёна

НИУ ВШЭ

20 февраля, 2020

План

1. flow-based generative models
2. architecture
   - split & squeeze
   - actnorm
   - invertible 1x1 convolution
   - affine coupling
3. experiments

Flow-based generative models

- exact latent-variable inference and log-likelihood evaluation
- efficient inference and efficient synthesis
- useful latent space for downstream tasks
- significant potential for memory savings

Dataset CelebA

Flow-based generative models

$$\mathbf{x} \sim p^*(\mathbf{x})$$

We want:

$$\mathbf{z} \sim p_\theta(\mathbf{z}),$$

$$\mathbf{x} = g_\theta(\mathbf{z}),$$

such as $\mathbf{z} = f_\theta(\mathbf{x}) = g_\theta^{-1}(\mathbf{x})$

Flow-based generative models

Reversible functions $\mathbf{f} = \mathbf{f}_1 \odot \mathbf{f}_2 \odot \cdots \odot \mathbf{f}_K$ Let $\mathbf{x} = \mathbf{h}_0, \mathbf{z} = \mathbf{h}_{K+1}$.
Then $\mathbf{h}_{i+1} = \mathbf{f}_{i+1}(\mathbf{h}_i)$

$$\log p_\theta(\mathbf{x}) = \log p_\theta(\mathbf{z}) + \log |\det \frac{d\mathbf{z}}{d\mathbf{x}}| =$$

$$= \log p_\theta(\mathbf{z}) + \sum_{i=1}^{K} \log |\det \frac{d\mathbf{h_i}}{d\mathbf{h_{i-1}}}|$$

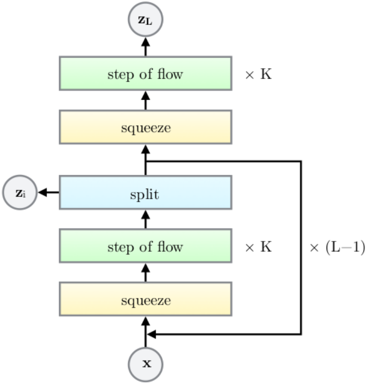
$\log |\det \frac{d\mathbf{h_i}}{d\mathbf{h_{i-1}}}|$ – logarithm of the absolute value of the determinant of the Jacobian matrix $\frac{d\mathbf{h_i}}{d\mathbf{h_{i-1}}}$

Flow-based generative models

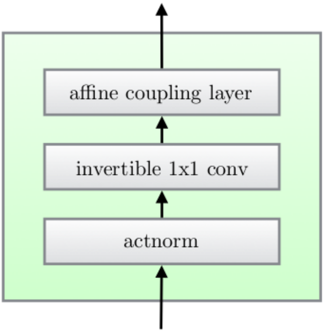
Jacobian matrix – upper triangluar matrix

$$\log|\det \frac{d\mathbf{h_i}}{d\mathbf{h_{i-1}}}| = \mathtt{sum}(\log|\mathtt{diag}(\frac{d\mathbf{h_i}}{d\mathbf{h_{i-1}}})|)$$

Architecture

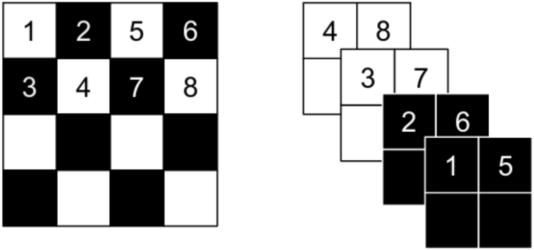


RealNVP



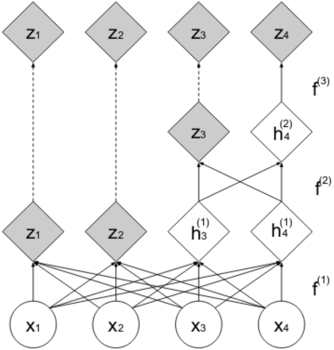
Step of Flow

Squeeze



Each $2 \times 2 \times c$ -> $1 \times 1 \times 4c$
From picture $s \times s \times c$ -> $\frac{s}{2} \times \frac{s}{2} \times 4c$

Split



Actnorm

Actnorm – activation normalization:

$$\forall i, j : \mathbf{y}_{i,j} = \mathbf{s} \odot \mathbf{x}_{i,j} + \mathbf{b}$$

Reverse function:

$$\forall i, j : \mathbf{x}_{i,j} = (\mathbf{y}_{i,j} - \mathbf{b}) / \mathbf{s}$$

Log-determinant:

$$h \cdot w \cdot \mathrm{sum}(\log |\mathbf{s}|)$$

initial $(\mathbf{s}, \mathbf{b})$: post-actnorm activations per-channel have zero mean and unit variance

Invertible 1x1 conv

Weight matrix – rotation matrix

$$\forall i, j : \mathbf{y}_{i,j} = \mathbf{W}\mathbf{x}_{i,j},$$

Обратная функция:

$$\forall i, j : \mathbf{x}_{i,j} = \mathbf{W}^{-1}\mathbf{y}_{i,j}$$

Log-determinant:

$$\log \left|\mathtt{det}\left(\frac{d\mathtt{conv2D}(\mathbf{h}, \mathbf{W})}{d\mathbf{h}}\right)\right| = h \cdot w \cdot \log \left|\mathtt{det}(\mathbf{W})\right|$$

W initialized as a random rotation matrix with log-determinant=0

LU-decomposition of Invertible 1x1 Conv

$$\mathbf{W} = \mathbf{PL}(\mathbf{U} + \mathtt{diag}(\mathbf{s}))$$

P - permutation matrix
L - lower triangular matrix with ones on the diagonal
U - lower triangular matrix with zeros on the diagonal
Log-determinant:

$$\log |\mathtt{det}(\mathbf{W})| = \mathtt{sum}(\log |\mathbf{s}|)$$

Affine coupling layer

Function:

$$\mathbf{x}_a, \mathbf{x}_b = \texttt{split}(\mathbf{x})$$

$$(\log \mathbf{s}, \mathbf{t}) = \texttt{NN}(\mathbf{x}_b)$$

$$\mathbf{s} = \exp \log \mathbf{s}$$

$$\mathbf{y}_a = \mathbf{s} \odot \mathbf{x}_a + \mathbf{t}$$

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

$$\mathbf{y} = \texttt{concat}(\mathbf{y}_a, \mathbf{y}_b),$$

Reverse Function:

$$\mathbf{y}_a, \mathbf{y}_b = \texttt{split}(\mathbf{y})$$

$$(\log \mathbf{s}, \mathbf{t}) = \texttt{NN}(\mathbf{y}_b)$$

$$\mathbf{s} = \exp \log \mathbf{s}$$

$$\mathbf{x}_a = (\mathbf{y}_a - \mathbf{t})/\mathbf{s}$$

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

$$\mathbf{x} = \texttt{concat}(\mathbf{x}_a, \mathbf{x}_b),$$

Log-determinant:

$$\mathrm{sum}(\log(|\mathbf{s}|)$$

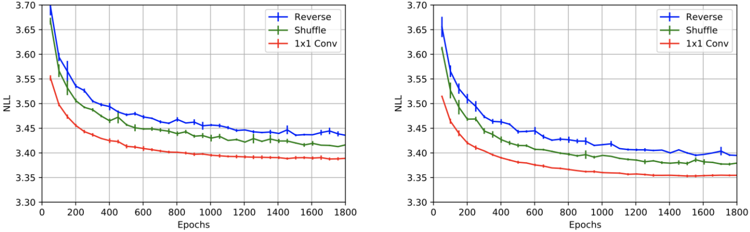
Affine coupling layer

- `NN` - nonlinear mapping
- `split` splits the input tensor into two halves along the channel dimension, `concat` - reverse operation.

additive coupling layer - special case with $s = 1$ and a log-determinant of 0.

Results

Experiments



(a) Additive coupling.

(b) Affine coupling.

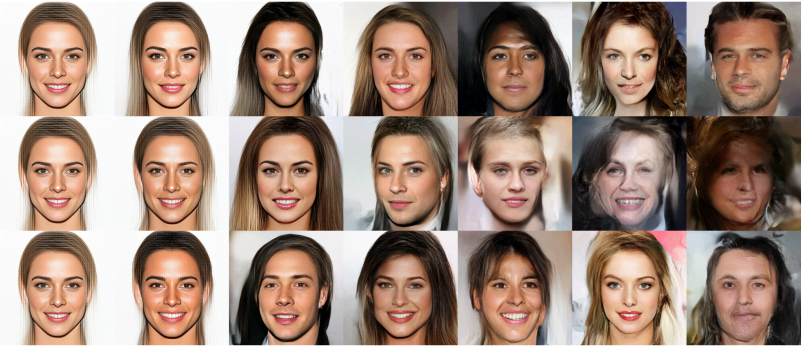
Table 2: Best results in bits per dimension of our model compared to RealNVP.

Model	CIFAR-10	ImageNet 32x32	ImageNet 64x64	LSUN (bedroom)	LSUN (tower)	LSUN (church outdoor)
RealNVP	3.49	4.28	3.98	2.72	2.81	3.08
Glow	**3.35**	**4.09**	**3.81**	**2.38**	**2.46**	**2.67**

Experiments

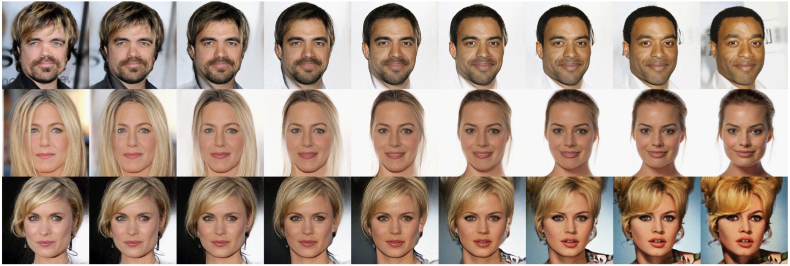
Left: L=4, Right: L=6

Experiments



Temperatures: 0, 0.25, 0.6, 0.7, 0.8, 0.9, 1.0

Experiments



Experiments

(a) Smiling

(b) Pale Skin

(c) Blond Hair

(d) Narrow Eyes

(e) Young

(f) Male

Questions

1. Что такое log-determinant? Запишите формулу.
2. Какие три основных слоя предлагаются авторами статьи «Glow: Generative Flow with Invertible 11 Convolutions»? Кратко опишите их.
3. Как авторы оптимизировали вычисление Invertible 1x1 Convolution?

References

[1] Diederik P. Kingma & Prafulla Dhariwal (2018). Glow: Generative Flow with Invertible 1x1 Convolutions *arXiv preprint arXiv:1807.03039.*

[2] Dinh, L., Krueger, D., & Bengio, Y. (2014). Nice: non-linear independent components estimation. *arXiv preprint arXiv:1410.8516.*

[3] Dinh, L., Sohl-Dickstein, J., & Bengio, S. (2016). Density estimation using Real NVP. *arXiv preprint arXiv:1605.08803.*