Neural ODE. FFJORD.

Gontar Daniil,
Faculty of CS, HSE
2019

Plan:

- Neural ODE
- Neural ODE in generative models: FFJORD
- FFJORD Perfomance
- Conclusion

Neural ODE

ResNet with skip-connections: $z(t+1) = z(t) + f_t(z(t), \theta)$

Similar to ODE! $\frac{dz(t)}{dt} = f(z(t), t, \theta)$

$t \in [0, T], z(0) = x$

$f(z(t), t, \theta)$ is a neural network

Continuous Backprop

Adjoint function: $a(t) = \frac{\partial L}{\partial z(t)}$

Math magic: $\frac{\partial L}{\partial \theta} = \int_0^T a(t) \frac{\partial}{\partial \theta} f(z(t), t, \theta) dt$

That is continuous backprop

Mode Collapse(review)

Normalizing Flows

$$z(0) \sim p_0(z(0))$$

$$z(T) = f(z(0), \theta)$$

$$z(T) \sim p_T(z(T))$$

$$\log p_T(z(T)) = \log p_0(z(0)) - \sum_{t=1}^{T} J_t$$

J - is the determinant of Jacobian

Continuous NFs

That is an ODE for log-likelihood on previous slide

$$\frac{d \log p_t(z(t))}{dt} = -tr\left(\frac{\partial f(z(t), t, \theta)}{\partial z(t)}\right)$$

Trace evaluation is much cheaper than determinant evaluation!!!

FFJORD extras

Trace tricks:

$$\mathrm{Tr}(A) = E_{p(\boldsymbol{\epsilon})}[\boldsymbol{\epsilon}^T A \boldsymbol{\epsilon}]$$

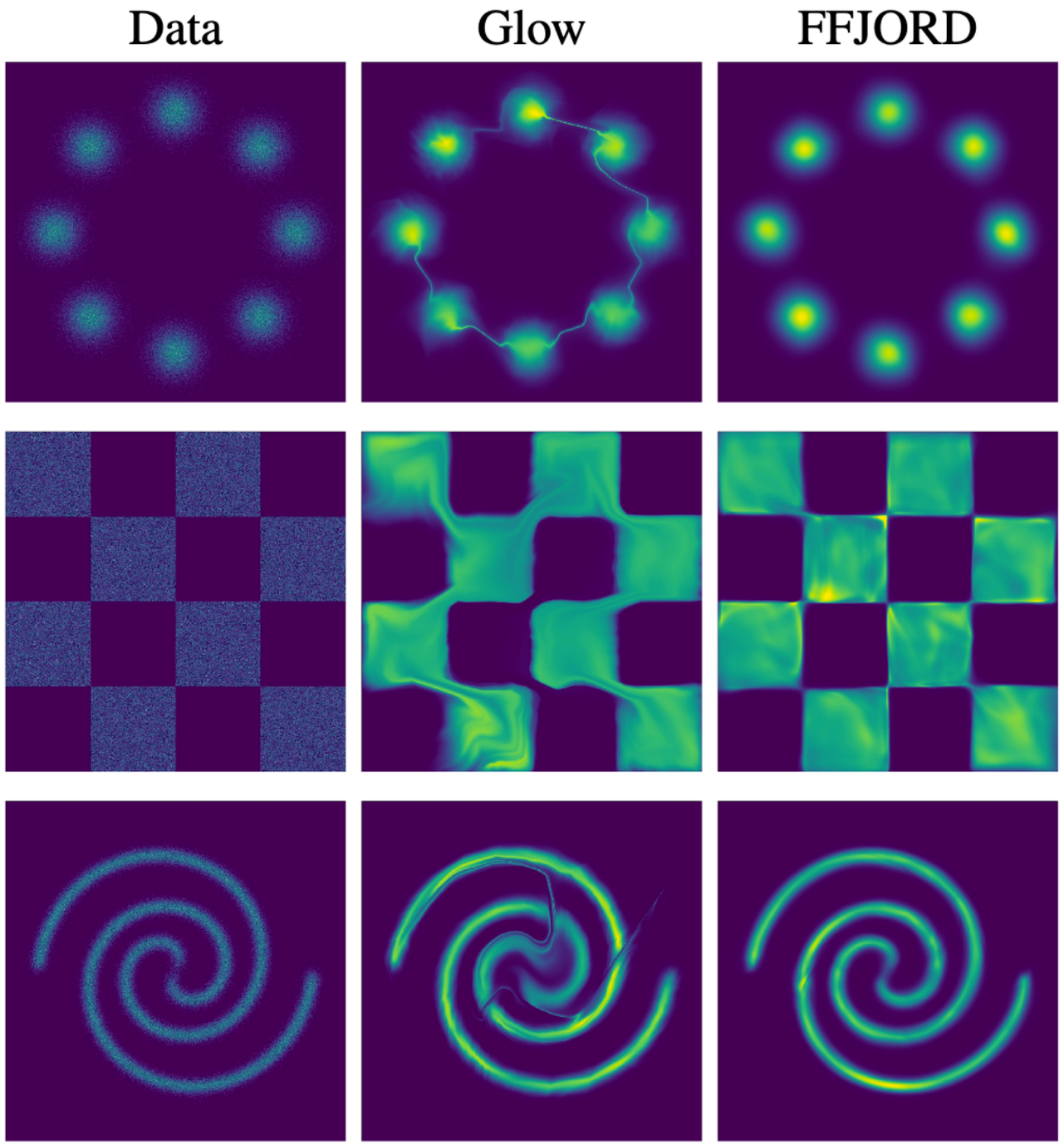
$$\mathrm{Tr}\underbrace{\left(\frac{\partial f}{\partial \mathbf{z}}\right)}_{D\times D} = \mathrm{Tr}\underbrace{\left(\frac{\partial g}{\partial h}\frac{\partial h}{\partial \mathbf{z}}\right)}_{D\times D} = \mathrm{Tr}\underbrace{\left(\frac{\partial h}{\partial \mathbf{z}}\frac{\partial g}{\partial h}\right)}_{H\times H} = \mathbb{E}_{p(\boldsymbol{\epsilon})}\left[\boldsymbol{\epsilon}^T \frac{\partial h}{\partial \mathbf{z}}\frac{\partial g}{\partial h}\boldsymbol{\epsilon}\right]$$

FFJORD Perfomance: Time complexity

- Evaluate f: $O(DH)$
- NF models: $O((DH + D^3)L)$
- CNF models: $O((DH + D^2)L')$
- FFJORD: $O((DH + D)L')$

Where: D - dim of data
H - size of the largest hidden layer
L - number of transformations
L’ - number of f evaluations used by ODE solver

FFJORD Perfomance: Density estimation



Summary

- Neural ODEs - new trend in ML
- Normalizing flows - useful instrument for density estimation
- Continuous NF = Neural ODEs + NF
- Continuous NF - faster and more powerful than NF
- FFJORD = CNF + trace tricks
- FFJORD - one of the fastest models, that use CNF

Useful materials:

- https://arxiv.org/pdf/1810.01367.pdf (FFJORD article)
- https://www.youtube.com/watch?v=8yJekeeGp_I (D.P.Vetrov: Neural ODEs)