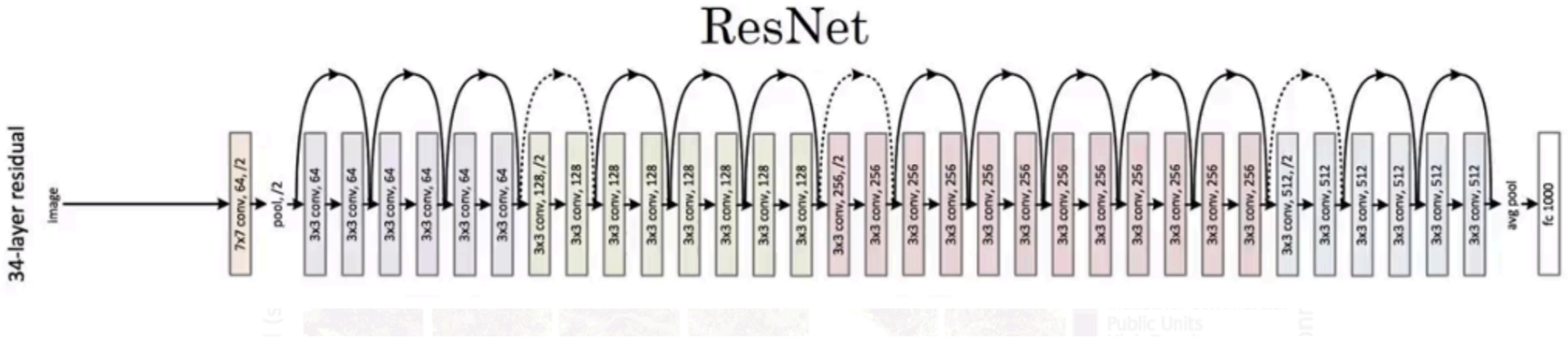
Neural ODE.

Gontar Daniil,
Faculty of CS, HSE
2019

Plan:

- ResNet: Old but Gold. What is Neural ODE?
- How to back-prop?
- Advantages of Neural ODE
- Normalizing flows and Neural ODE
- Conclusion

ResNet: Recap



What is the key advantage of ResNet?

(Ans: No vanishing gradients)

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

ResNet and Neural ODE

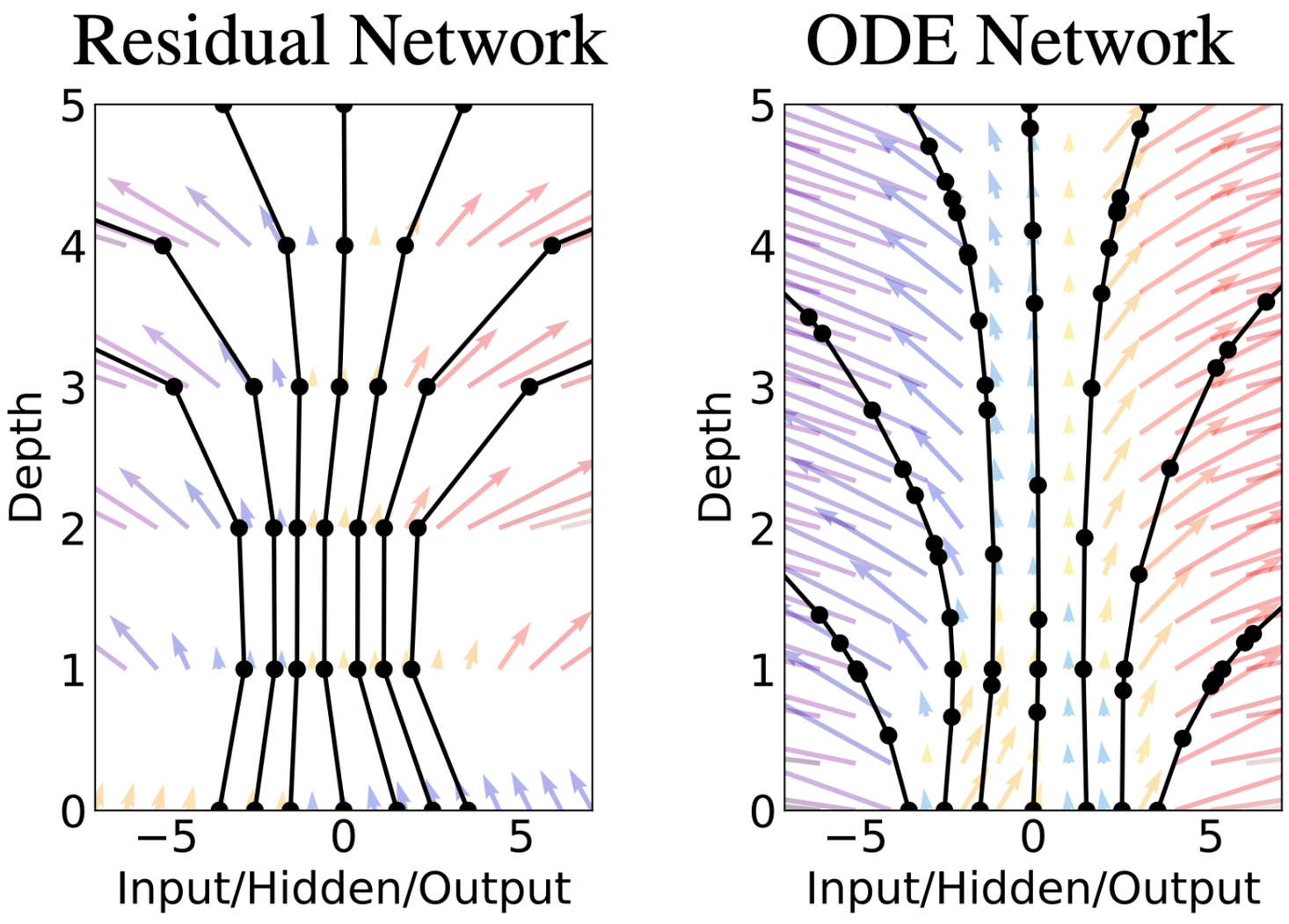


Figure 1: *Left:* A Residual network defines a discrete sequence of finite transformations. *Right:* A ODE network defines a vector field, which continuously transforms the state. *Both:* Circles represent evaluation locations.

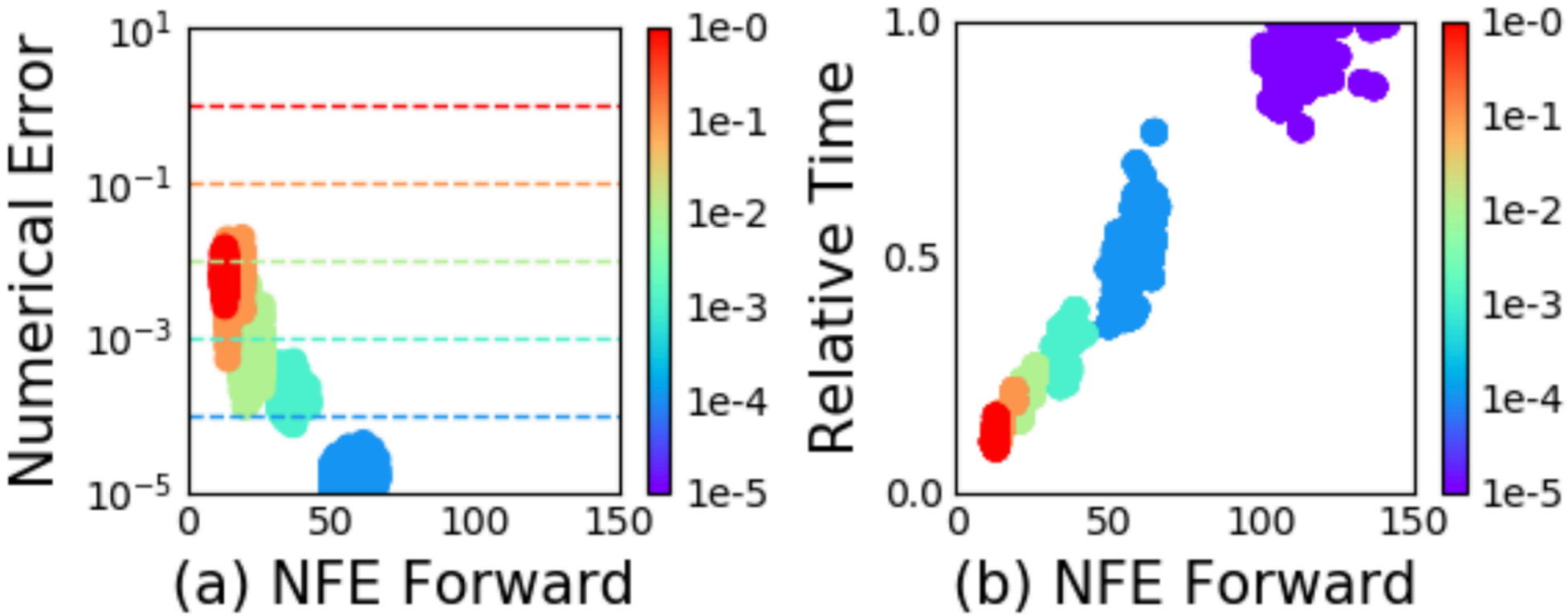
Continuous Backprop

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

Adaptive computation time



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

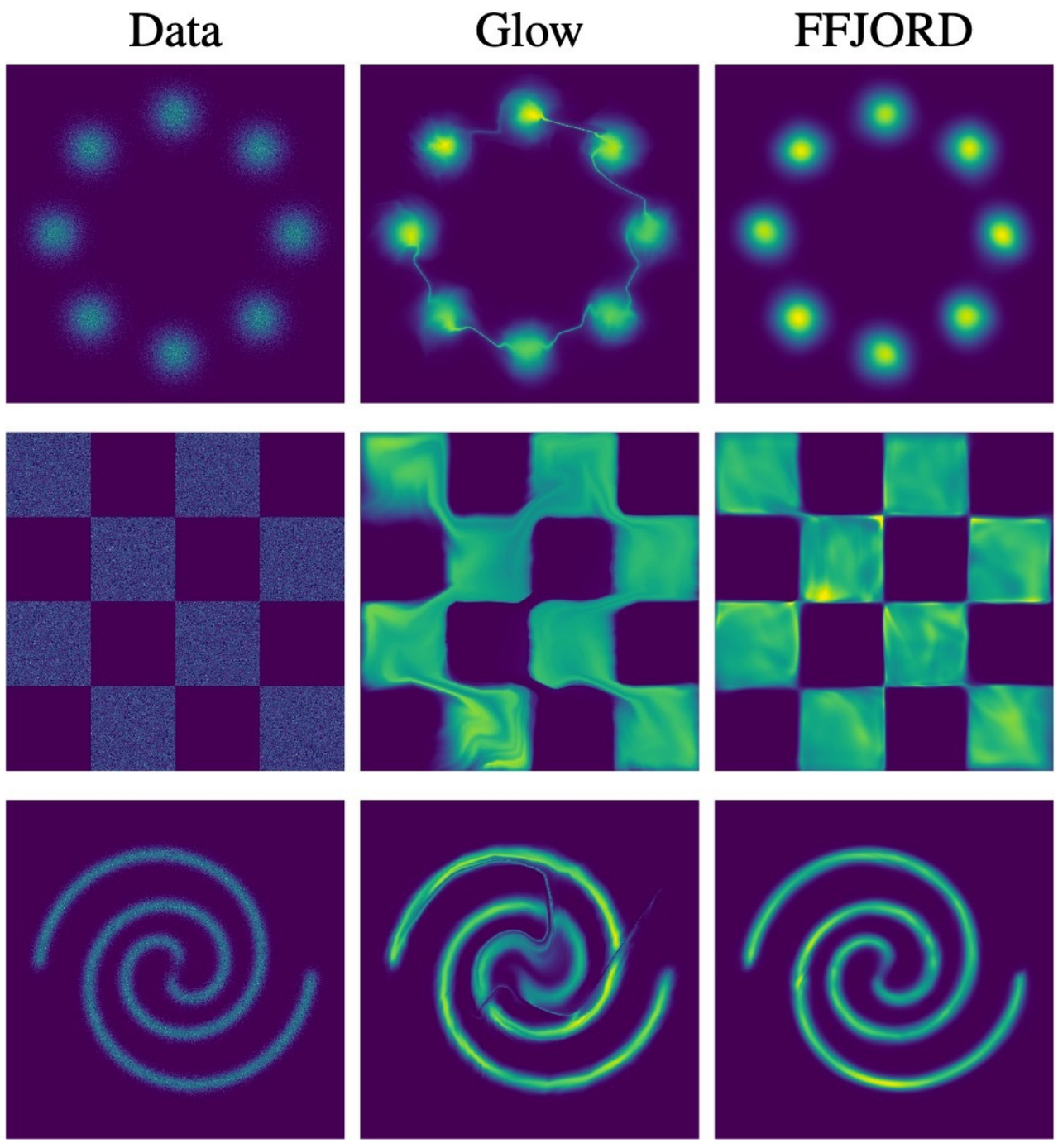
Continuous NFs

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!!!

Continuous NF Perfomance: toy density estimation



Summary

- Neural ODE - amazing idea from old ResNet
- Neural ODEs have adaptive computation time
- Ability to use more powerful numeric techniques, than Euler’s scheme
- CNF - powerful technique based on Neural ODEs for density estimation

Useful Links

- Neural ODE paper: https://arxiv.org/pdf/1806.07366.pdf
- Dmitry Vetrov's lecture on Neural ODE: https://www.youtube.com/watch?v=8yJekeeGp_I
- FFJORD: https://arxiv.org/abs/1810.01367

Вопросы для самостоятельной

- Запишите дифференциальное уравнение, численное решение схемой Эйлера которого задаёт ResNet.
- Какие преимущества по сравнению с ResNet есть у нейродиффуров?
- Для чего используются нормализующие потоки? В чем преимущество непрерывных норм потоков?