Structure-Preserving Deep Learning

Matthias J. Ehrhardt

Department of Mathematical Sciences, University of Bath, UK

The Leverhulme Trust

UKRI Engineering and Physical Sciences Research Council

m4DL

Notation: Neural Network

Define **neural network** $\Phi_\theta : X \to Y$ recursively: $\Phi_\theta(x) = z^K$

$$z^0 = x \in X$$
$$z^{k+1} = f^k(z^k, \theta^k), \quad k = 0, \dots, K-1$$

with generic **layers**

$$f^k : Z^k \times \Theta^k \to Z^{k+1}, \quad k = 0, \dots, K-1$$

- Classical, fully-connected layer defined by

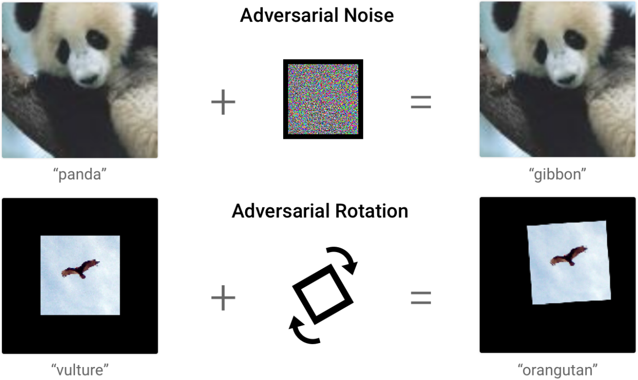
  $$\begin{aligned} f : \mathbb{R}^M \times (\mathbb{R}^{M' \times M} \times \mathbb{R}^{M'}) &\to \mathbb{R}^{M'} \\ (z, (A, b)) &\mapsto \sigma(Az + b), \end{aligned}$$

  where σ is an element-wise nonlinearity (ReLU, tanh etc.)
- A is often replaced by a convolutional operator
- **Training goal**: dataset $\{(x_n, y_n)\}_n$

  $$\min_{\theta \in \Theta} \frac{1}{N} \sum_{n=1}^{N} L(\Phi_\theta(x_n), y_n) + R(\theta)$$

Deep Learning and Robustness

- Deep learning often is not robust (e.g. noise, rotations, ...)



https://ai.googleblog.com/2018/09/introducing-unrestricted-adversarial.html

- Data augmentation ...
- **Our approach**: Design deep learning architectures with mathematical guarantees (e.g. stability, equivariance, invertibility, manifolds, ...)

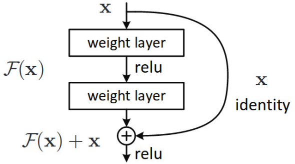
Residual networks as discretised ODEs

- "Standard" Neural Networks

$$z^{k+1} = \sigma(A^k z^k + b^k)$$

- Deep Residual Neural Networks (ResNet) He, Zhang, Ren, Sun 2015 ($>$ 85000 citations on GoogleScholar)

$$z^{k+1} = z^k + \Delta t\, \sigma(A^k z^k + b^k)$$



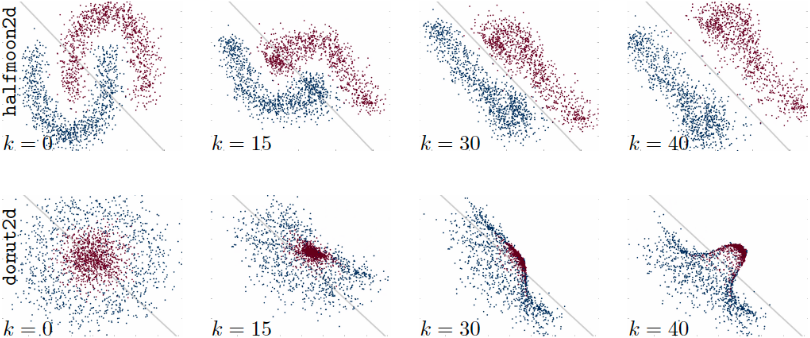
ResNet is Forward Euler discretization $\dot{z}(t) \approx \frac{z(t+\Delta t) - z(t)}{\Delta t}$ of

$$\dot{z}(t) = \sigma(A(t)z(t) + b(t)), \quad t \in [0, T]$$

with continuous-time mappings A, b. $z^k := z(k\Delta t)$ ...

Haber and Ruthotto 2018; Li et al. 2018, Benning et al. 2019

ResNet in action



Interpretation as discrete optimal control

The **deep learning problem** can be seen as the discretization of

Optimal control problem

$$\min_{\theta} \frac{1}{N} \sum_{n=1}^{N} L(z_n(T), y_n)$$

subject to

$$\dot{z}_n = f(z_n, \theta), \quad z_n(0) = x_n \in X.$$

Why is the optimal control point of view useful:

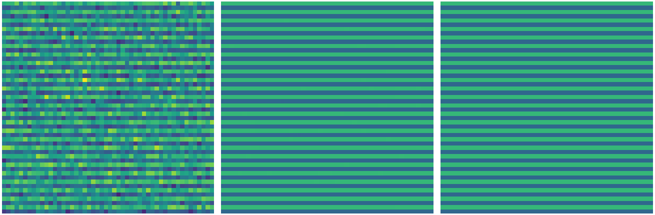
- it states the deep learning problem in two lines
- can be used to create new architectures
- continuous models are useful simplifications of reality, amenable for analysis
- what ODE properties carry over to discrete neural networks?

Haber and Ruthotto 2018; Li et al. 2018, Benning et al. 2019

What happens when images are rotated?

$$\Phi(y) = x$$

Training data

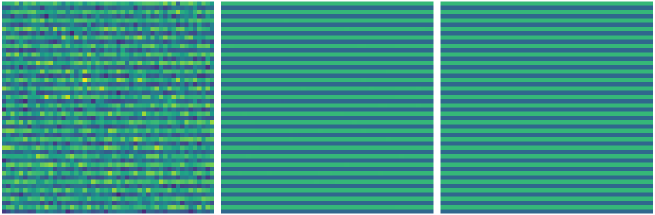


Noisy Ordinary Equivariant

What happens when images are rotated?

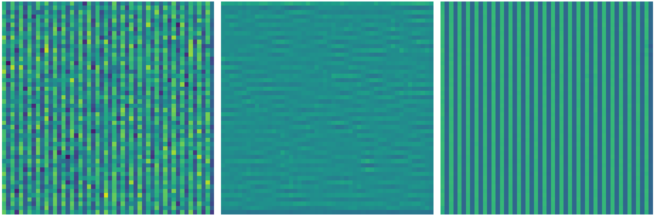
$$\Phi(y) = x$$

Training data



Noisy Ordinary Equivariant

Test data

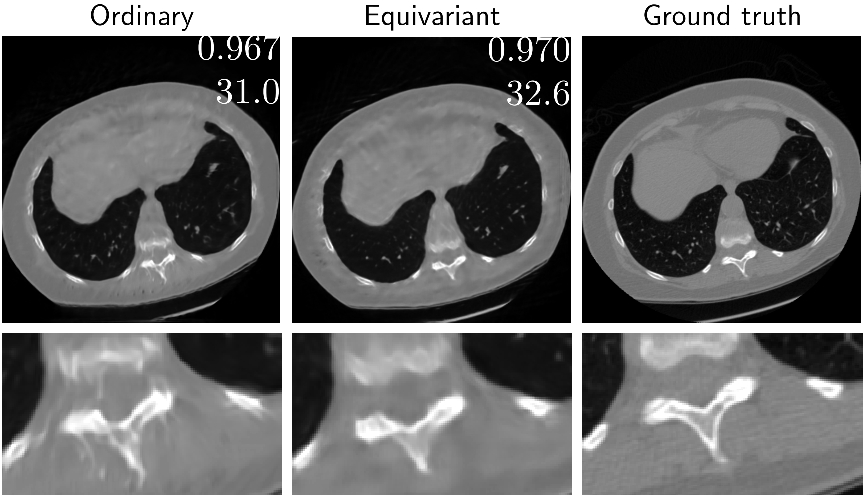


Noisy Ordinary Equivariant

CT Results

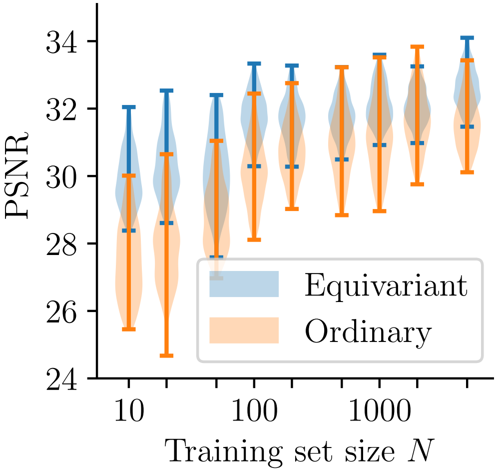
Equivariant improves upon Ordinary:

- **higher** SSIM and PSNR
- **fewer** artefacts and **finer** details



CT Results

- Equivariant improves upon Ordinary on **small** training sets



Take Away Messages

- **Continuum modelling** of neural networks opens the toolbox of mathematical and numerical analysis
- **Connections** of deep learning to ODEs, optimal control, group theory ...
- Design of neural networks with certain **structure**: stability, equivariance, invertibility, manifolds, ...

E. Celledoni, M. J. Ehrhardt, C. Etmann, R.I. McLachlan, B. Owren, C. B. Schönlieb, F. Sherry, *Structure-Preserving Deep Learning*, EJAM 2021

E. Celledoni, M. J. Ehrhardt, C. Etmann, B. Owren, C. B. Schönlieb, F. Sherry, *quivariant Neural Networks for Inverse Problems*, Inverse Problems 2021