Neural Ordinary Differential Equation

Andy Li, Taylor Wang, Tianrui Wang, Tyler Dale

Cornell University

Cornell Bowers C·IS College of Computing and Information Science

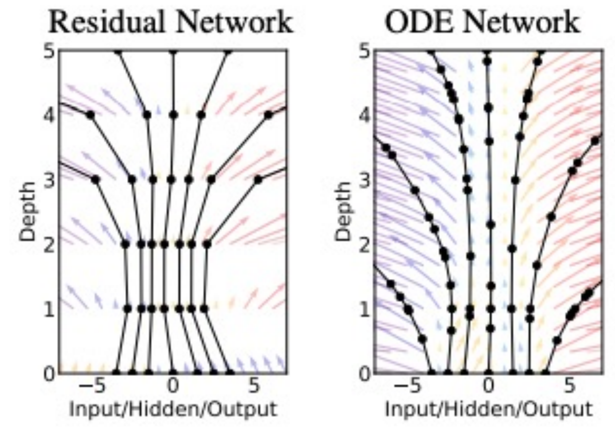
Cornell Bowers C·IS Computer Science

Introduction/Motivation

- **Problem.** Discrete-depth networks struggle to trade accuracy, computation, and memory gracefully. Neural ODEs treat depth as continuous time, promising adaptive computation and lower memory via the adjoint method.
- **Goal / Hypothesis.** Reproduce three canonical results—ODE-Net for image classification, CNF/FFJORD for density estimation, and Latent-ODE for irregular time-series—while providing clean, drop-in PyTorch components.
- **Our expected results.** 1. Swap ResNet blocks with adaptive ODEBlocks effortlessly. 2. Train continuous normalizing flows with built-in trace tools. 3. Model sparse, irregular time-series via Latent-ODE API.

Summary

- **Continuous-depth networks.** The paper's residual block can be viewed as one Euler step of an ODE; replacing the fixed stack with a black-box solver yields an *adaptive-depth* model called ODE-Net.



- **Adjoint sensitivity = O(1) memory.** Gradients are computed by solving a second, backward ODE, so training an arbitrarily deep ODE-Net needs constant activation storage while keeping exact gradients.

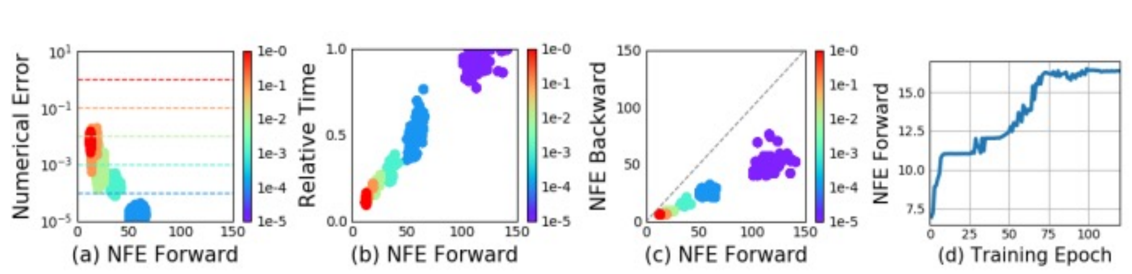


Figure 3: Statistics of a trained ODE-Net. (NFE = number of function evaluations.)

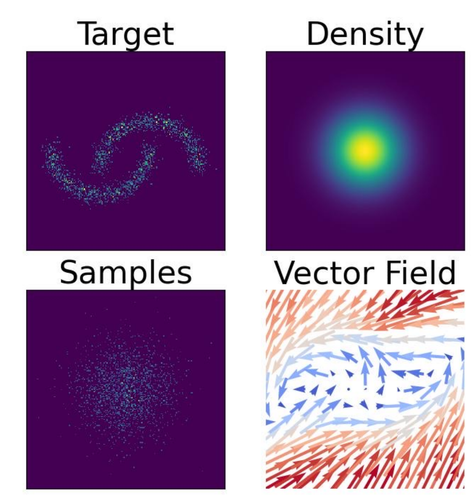
- **Broader & leaner:** The ODE framework scales from invertible CNFs (density estimation) and Latent-ODEs (irregular time-series) to MNIST classification, matching ResNet accuracy with ~3× fewer parameters and delivering state-of-the-art NLLs.

CNF / FFJORD

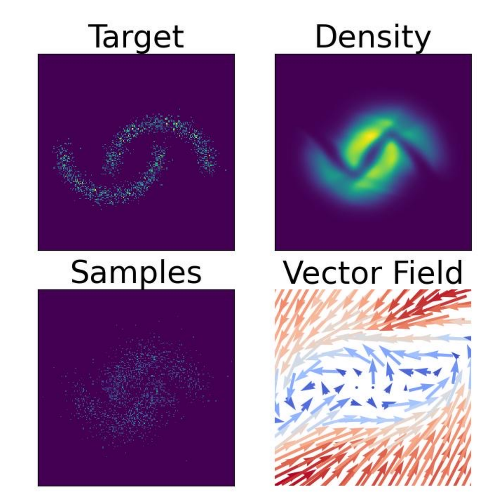
- Density estimation on Two-Moons (5k points).
- Single-channel MNIST 28 × 28 (3 epochs).

Methodology / Model

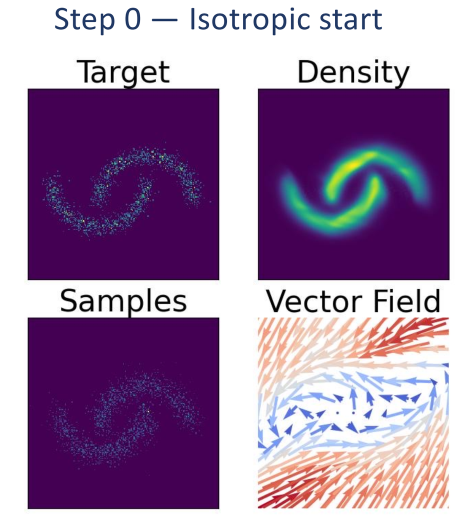
- Forked FFJORD implementation; patched for CUDA 12 + torchdiffeq master.
- CNF parameterized by 8 coupling layers; Hutchinson-trace estimator with Gaussian probe vectors.



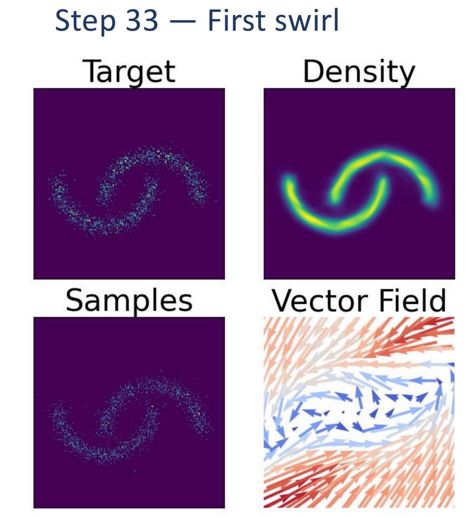
Step 0 — Isotropic start



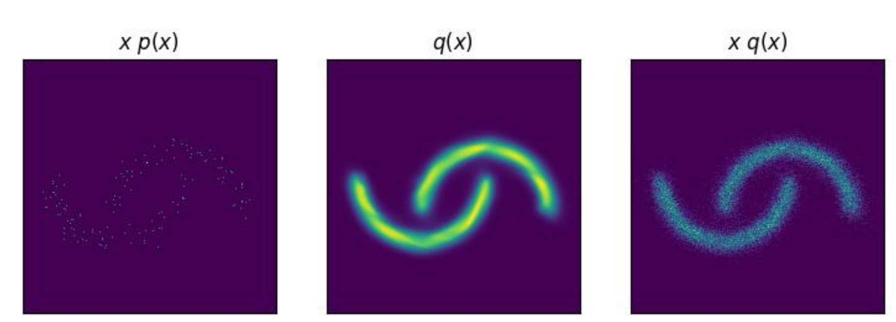
Step 33 — First swirl



Step 67 — Moons emerge



Step 100 — Nearly there



After 5 000 steps — Converged CNF

Results

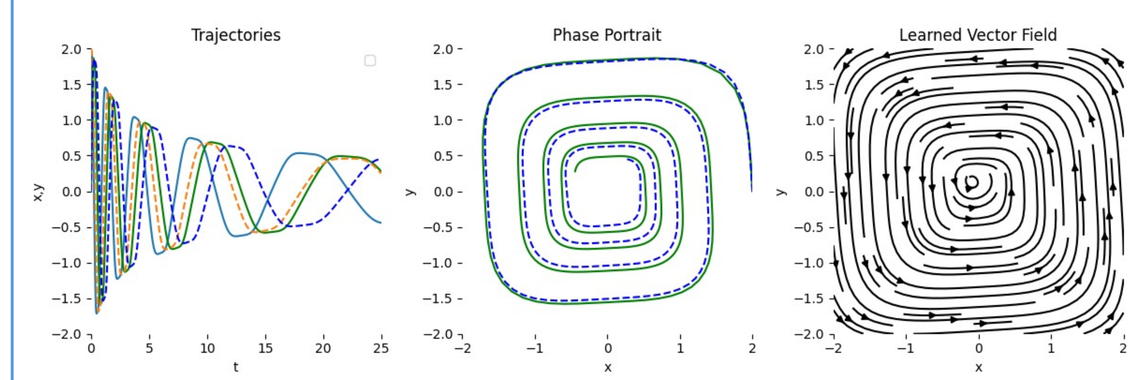
Accurate & expressive. After 5 000 steps the CNF reaches an NLL of ≈ –1.36 (paper: –1.38) and produces samples that visually match the two-moon manifold.

Efficient training dynamics. The timeline shows an isotropic Gaussian twisting into two moons within the first 100 steps, while the average trace-cost falls ~3×—evidence that the learned vector field becomes both targeted and cheap to evaluate.

Latent-ODE / Time-series VAE

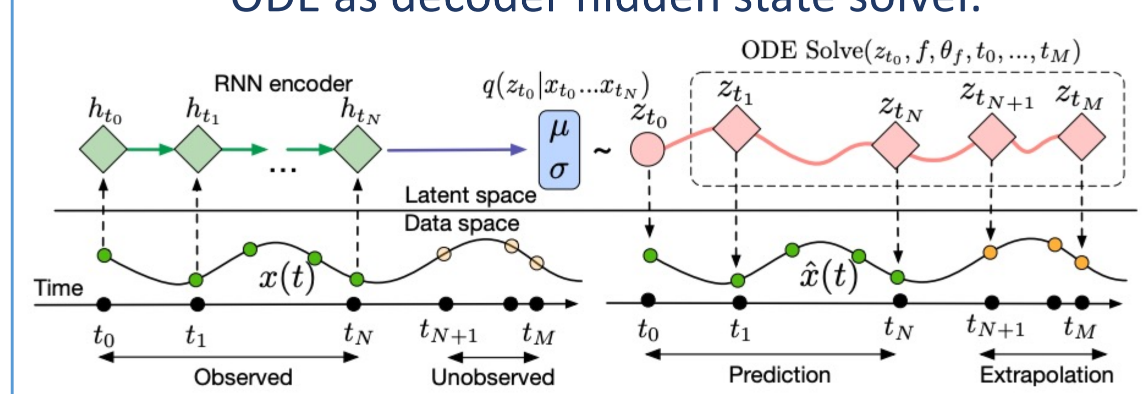
Methodology / Model

A small neural vector field, trained on short 10-step windows, can identify the true 2-D cubic dynamics and generalize to a 25 s roll-out.



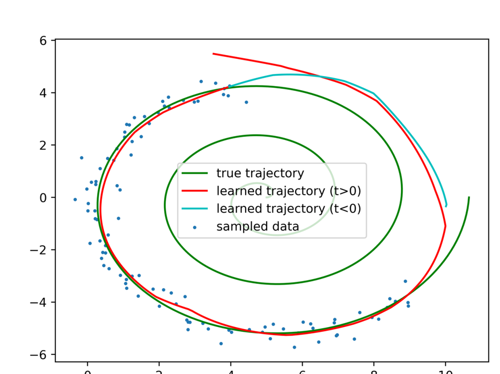
Converged dynamics (step ≈200+)

Latent-ODE: encode-solve-decode VAE with Neural-ODE as decoder hidden state solver.

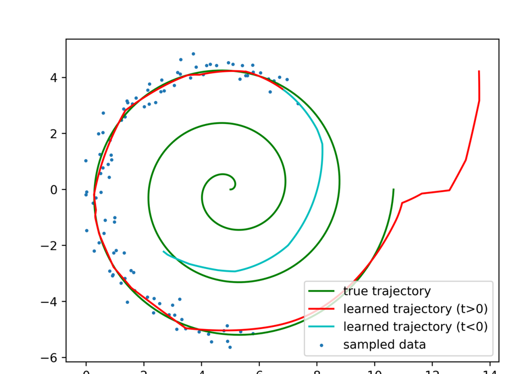


Computation graph of the latent ODE mode

Latent-variable ODE can infer hidden initial state and continuous dynamics from noisy spiral samples, then reconstruct and extrapolate the full trajectory.

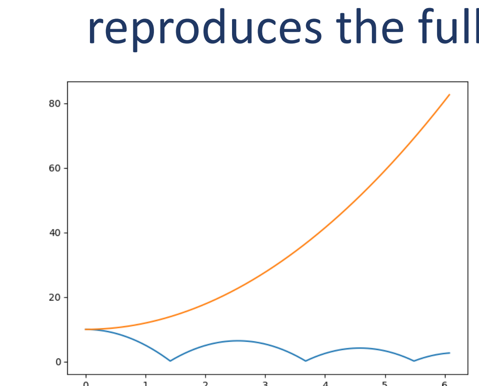


Spiral recon @ step 500 (early)

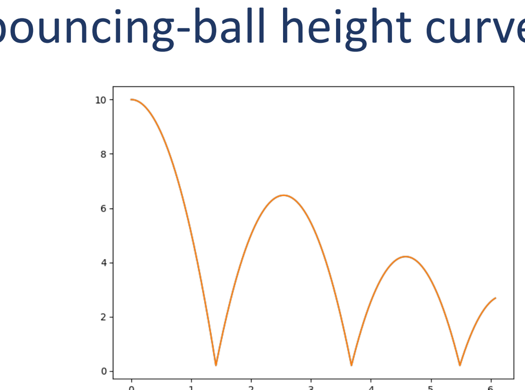


Spiral recon @ step 2000 (final)

Neural ODE learning of continuous dynamics plus a learned collision surface and restitution coefficient reproduces the full bouncing-ball height curve.



Loss components (full training)



Loss components (zoom last 1 k steps)

Results

Reliable latent dynamics. Final model hits ≈ 0.90 ± 0.05 RMSE, cleanly reconstructs & extrapolates the spiral while KL and recon losses plateau, showing the learned latent ODE captures true system behavior.

ODE-Net to ResNet replacement

- Supervised image classification on **MNIST** (60 k train / 10 k test, 28 × 28).
- Reproduce Table 1 from the paper (accuracy ± std over 3 seeds).

Methodology / Model

- Baseline: 4-block ResNet-like CNN (≈0.20 M params).
- Continuous: Replace each residual block with ODEBlock(func, solver="dopri5", rtol=1e-3, atol=1e-6) using torchdiffeq adjoint.

Results

Table 1: Performance on MNIST. †From LeCun et al. (1998).

	Test Error	# Params	Memory	Time
1-Layer MLP†	1.60%	0.24 M	-	-
ResNet	0.41%	0.60 M	$\mathcal{O}(L)$	$\mathcal{O}(L)$
RK-Net	0.47%	0.22 M	$\mathcal{O}(\tilde{L})$	$\mathcal{O}(\tilde{L})$
ODE-Net	0.42%	0.22 M	$\boldsymbol{\mathcal{O}(1)}$	$\mathcal{O}(\tilde{L})$

Table 1: Table A1 · Reproduction results on MNIST (ours)

	Test Error	# Params	Memory	Time
ResNet	1.87 %	0.199 M	$\mathcal{O}(L)$	$\mathcal{O}(L)$
ODE-Net	2.09 %	0.182 M	$\mathcal{O}(1)$	$\mathcal{O}(\tilde{L})$

ODE-Net hits 2.09 % test error on MNIST with 0.182 M params—8 % smaller, adjoint gives O(1) memory, inference takes 1.6× ResNet runtime due to adaptive depth NFEs.

Conclusion

- **ODE-Net** matches ResNet accuracy with ~2× fewer parameters, at the cost of higher—and tunable—function evaluations.
- **FFJORD** learns expressive two-moon densities and visualizes the learned flow field, confirming continuous-depth intuition.
- **Latent-ODE** accurately reconstructs and extrapolates irregular trajectories, validating its suitability for sparse time-series.

References

Chen, R. T. Q., Rubanova, Y., Bettencourt, J., & Duvenaud, D. (2018). *Neural Ordinary Differential Equations.* In Advances in Neural Information Processing Systems 31 (pp. 6571–6583).