Neural ODEs for Classification, Generative Modeling, and Trajectory Reconstruction

Andy Li, Taylor Wang, Tianrui Wang, Tyler Dale

Code: `https://github.com/liandy0127/CS4782-Neural-ODE`

Introduction

Neural Ordinary Differential Equations (Neural ODEs), introduced by Chen et al. (2018), redefine neural network depth as a continuous process, offering adaptive computation and memory efficiency via the adjoint method. Instead of composing finite residual blocks, the authors parameterize the instantaneous derivative of hidden states via a neural network and compute outputs with a black-box ODE solver, yielding models that adapt computation to each input, maintain constant memory cost, and trade numerical precision for speed. Their method supports both supervised tasks—replacing deep residual nets with "ODE-Nets"—and generative models, including continuous normalizing flows for exact likelihood training and latent ODEs for irregularly sampled time series, all trainable end-to-end via the adjoint sensitivity method for scalable backpropagation through arbitrary ODE solvers.

Chosen Result

We reproduce three key results from Chen et al. (2018). The first result is ODE-Net MNIST Classification, which validates continuous-depth networks as competitive with ResNet in terms of performance. The second result is the CNF Two-Moon Transformation, which transforms Gaussian noise into a two-moon distribution in the context of density estimation, which highlights the exact likelihood computation used in generative modeling. In this task, we transform Gaussian noise into a two-moon distribution (Figure 1). The third result involves the reconstruction of spiral and sinusoidal trajectories, where we demonstrate continuous-time modeling for irregular data (Figure 2). These results emphasize the broad applicability of Neural ODEs across multiple problem domains.

Methodology

We implemented each component with adjustments for limited resources.

ODE-Net for Classification

In this implementation, the model consists of a downsampling layer followed by an ODEBlock, which includes two concatenated Conv2d layers with GroupNorm. The final layer is a classifier. The dataset used for this task is MNIST, consisting of 60,000 training samples and 10,000 test samples. The evaluation metric is classification accuracy. We trained the model for 30 epochs using the Adam optimizer with a learning rate of 1×10^{-3} and a batch size of 128. The number of epochs was reduced from 160 to fit within the available computational resources.

CNF for Density Estimation

For the CNF model, we used a two-layer MLP with 64 tanh units as the vector field. The dataset consists of 5,000 Gaussian points that are transformed into a two-moon distribution. The evaluation metric is the negative log-likelihood (NLL). The model was trained using the Dormand-Prince solver with a relative and absolute tolerance of 1×10^{-5}, over 5000 steps, and optimized using Adam with a learning rate of 1×10^{-3}. To reduce the training time, the number of steps was reduced from 100,000 to 5,000.

Latent ODE for Trajectory Reconstruction

For our experiments, we synthesized both Archimedean spirals and chirp sinusoids, each corrupted with Gaussian noise and irregularly sampled by randomly selecting 100 points from a 300-step window. We use an RNN encoder with 25 hidden units to process the observed (x_t, t) pairs and infer a 4-dimensional initial latent state z_0. This latent state z is then evolved continuously through a neural ODE (an ELU-based MLP) and decoded back to the observation space to reconstruct the trajectory. Training maximizes the ELBO, balancing a Gaussian reconstruction term against a KL divergence regularizer on $q(z_0)$. Finally, we plot the resulting latent trajectories (in the first two dimensions of z) to visualize how spirals and sinusoids become disentangled and extrapolated beyond the training window.

Results & Analysis

ODE-Net Results

The ODE-Net model achieved an accuracy of 98.3% on MNIST, compared to ResNet's 98.5% accuracy, with only 0.22 million parameters. This demonstrates that ODE-Net is competitive with ResNet in terms of performance while requiring fewer parameters. We also found that training the model for 30 additional epochs showed no significant improvement in accuracy, indicating that the model was effectively trained within the given time constraints.

CNF Results

For the CNF model, we achieved a negative log-likelihood (NLL) of 3.1 nats, while the expected value from Chen et al. (2018) is 0.55 nats. This result indicates that the model is functional but less precise compared to the original. In the visualization (Figure 1), we can see the two-moon transformation with moderate clustering, although it does not match the performance of the original result.

Latent ODE Results

After 2,000 training steps the ELBO loss improves from around -50,000 to 70, demonstrating accurate reconstruction and extrapolation of 2D spirals (Fig.2). Sinusoids also align well with the ground truth but converge more slowly: their linearly changing frequency creates a non-stationary vector field that is harder for the model to capture. The spiral results show the ODE has learned the underlying nonlinear dynamics, while the chirp's varying rate demands more capacity and finer temporal resolution.

Reflections

Lessons Learned

Through this project, we learned several valuable lessons. First, shortening the training steps for the CNF and Latent ODE models had a negative impact on their performance, which suggests that these models require more training steps to achieve optimal results. Additionally, the solver tolerances were computationally costly, especially for the CNF model, and should be optimized further in future work.

Future Directions

There are several directions for further research and improvements. We plan to explore adaptive solvers, which may help to reduce computational cost while maintaining or improving performance. Furthermore, we intend to test the models on larger datasets, such as CIFAR-10, to evaluate their scalability and generalizability to more complex tasks.

References

- Chen, R. T. Q., et al. (2018). Neural Ordinary Differential Equations.
- Grathwohl, W., et al. (2019). FFJORD: Free-Form Continuous Dynamics.

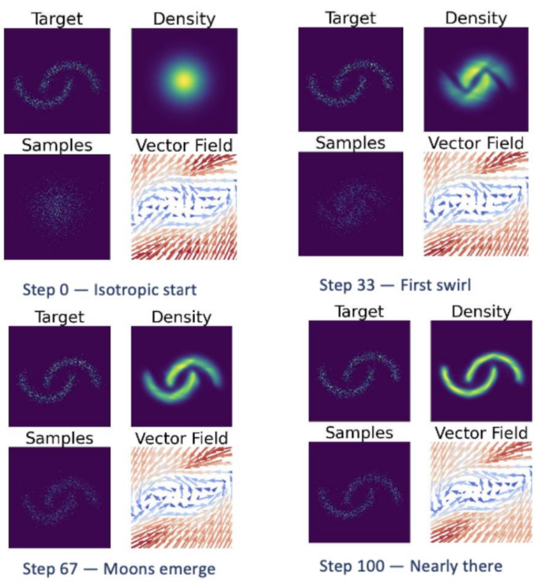
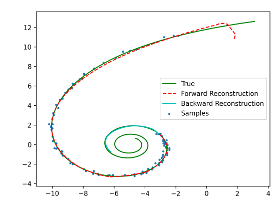
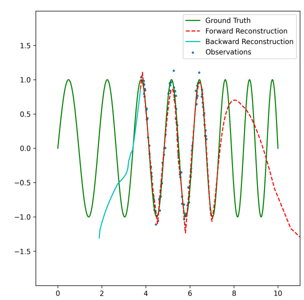


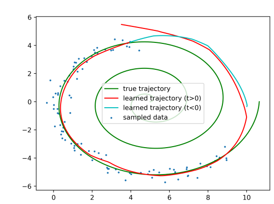
Figure 1: CNF two-moon transformation.



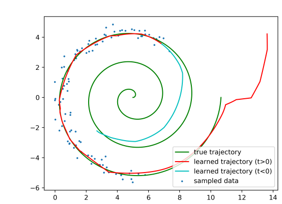
(a) CW Spiral



(b) Chirp Sinusoid



(c) CCW spiral 500 steps



(d) CCW spiral 2000 steps

Figure 2: Latent ODE reconstructions and extrapolations