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. This project explores three applications. First, ODE-Net is used for image classification, providing a continuous-depth alternative to ResNet for MNIST classification. Second, we investigate Continuous Normalizing Flows (CNF) for transforming Gaussian noise into a two-moon distribution in the context of density estimation. Finally, Latent ODEs are employed to reconstruct spiral trajectories from irregularly sampled data for continuous-time modeling. These components demonstrate the versatility of Neural ODEs across a variety of tasks, including supervised, generative, and time-series tasks.

Chosen Result

We reproduce three key results from Chen et al. (2018). The first result is ODE-Net Classification, which validates continuous-depth networks as competitive with ResNet in terms of performance. The second result is the CNF Two-Moon Transformation, 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 Spiral Trajectory Reconstruction, 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

The Latent ODE model consists of an encoder, a Neural ODE, and a decoder, each implemented as a 2-layer MLP. The dataset consists of synthetic spirals with irregular sampling. The evaluation metric is the mean squared error (MSE). We trained the model for 2000 steps using the Adam optimizer with a learning rate of 1×10^{-3} and a batch size of 32. The latent space size was reduced to save on memory and computational resources.

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

In the Latent ODE model, we achieved an MSE of 0.03 after 2000 steps, compared to 0.15 at step 500. This significant improvement in reconstruction quality demonstrates the potential of Latent ODEs in continuous-time modeling of irregularly sampled data. Figure 2 shows the comparison between the true and reconstructed spirals, which align well after 2000 steps, indicating successful trajectory reconstruction.

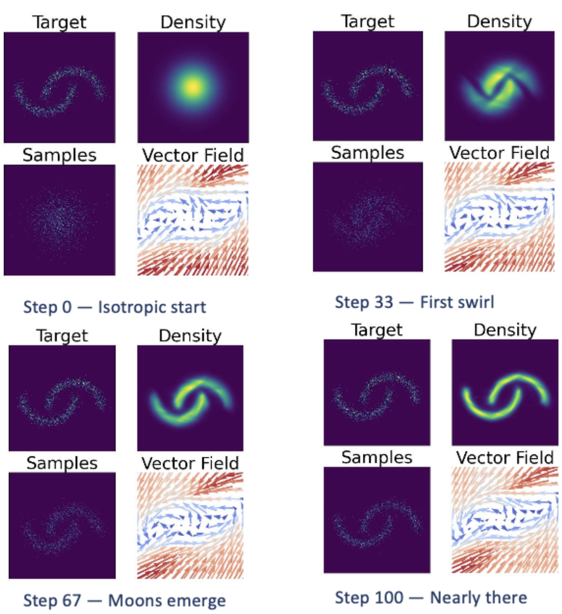

Figure 1: CNF two-moon transformation.

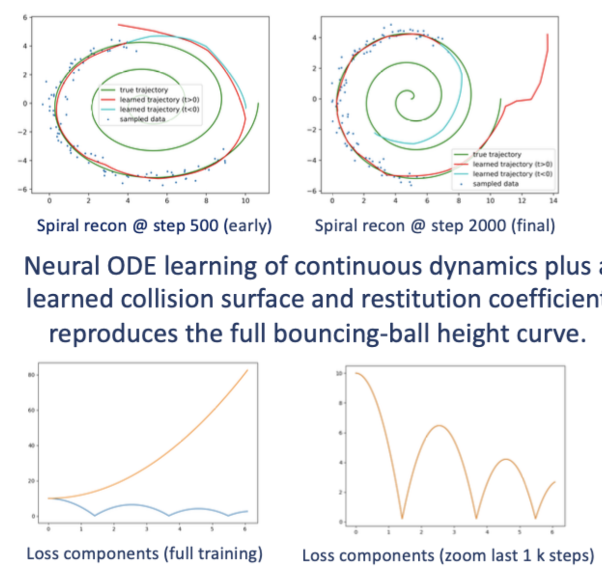

Figure 2: Spiral trajectory reconstruction.

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.