ICLR

Scalable Mechanistic Neural Networks

Jiale Chen (jiale.chen@ist.ac.at)

Dingling Yao

Adeel Pervez

Dan Alistarh

Francesco Locatello

ISTA Institute of Science and Technology Austria

Background: Mechanistic Neural Network (MNN)

MNN is a powerful tool.

	NeuralODE, UDE Chen et al. (2018) Rackauckas et al. (2020)	SINDy Brunton et al. (2016)	Neural Operators Li et al. (2020)	MNN Pervez et al. (2024)
Linear Discovery		✓		✓
Nonlinear Discovery				✓
Physical Parameters	✓	✓		✓
Forecasting	✓		✓	✓
Interpretability		✓		✓

Solvers

- Dense solver (dense matrices)
- Sparse solver (iterative methods – conjugate gradient)

Limitations

- High computational costs in time and memory
- Inefficient GPU utilization
- Sparse solver: numerical inaccuracies in large-scale problems

Complexities for sequence length T

	Time Complexity	Space Complexity
Dense Solver	$\boldsymbol{O(T^3)}$	$O(T^2)$
Sparse Solver	$O(T^2)$	$O(T^2)$

Adeel Pervez, Francesco Locatello, Stratis Gavves. Mechanistic neural networks for scientific machine learning. Proceedings of the 41st International Conference on Machine Learning, PMLR 235:40484-40501, 2024.

Scalable MNN (S-MNN)

Time and space complexities: $O(T)$

(S-)MNN = Encoders + Differentiable ODE Solver + Optional Decoders

- Encoders: learn dynamics from data
- Solver: solve linear systems for time sequence predictions
- Decoders: generate outputs for forecasting and parameter discovery

Three types of constraints

- Governing equation constraints
- Initial value constraints
- Smoothness constraints (different from the original MNN)

Forward pass (training & inference): solve linear system $\boldsymbol{A}\boldsymbol{y} = \boldsymbol{b}$ in the S-MNN solver

- Under all proper use cases, the linear system is over-constrained.
- The least squares solution is $\boldsymbol{y} = (\boldsymbol{A}^\top \boldsymbol{A})^{-1}\boldsymbol{A}^\top \boldsymbol{b}$, denoted as $\boldsymbol{y} = \boldsymbol{M}^{-1}\boldsymbol{\beta}$.
- $\boldsymbol{M}$ is a <u>banded symmetric matrix</u>, so $O(T)$ solver exists.

Backward pass (training only)

- $O(T)$ complexity: $\frac{\partial l}{\partial \boldsymbol{\beta}} = \boldsymbol{M}^{-1}\frac{\partial l}{\partial \boldsymbol{y}}$ and $\frac{\partial l}{\partial \boldsymbol{M}} = -\frac{\partial l}{\partial \boldsymbol{\beta}}\boldsymbol{y}^\top$

Input Sequence $x_1, x_2, \cdots, x_T$

Encoder Network

Coefficients c, d, Initial Values u, Time Intervals s

ODE Representation

Governing Equation Constraints (Same as MNN)

Linear ODE of Unknown 1D Function $y(t)$:

$$c_0(t)y(t) + c_1(t)\frac{\mathrm{d}y(t)}{\mathrm{d}t} + c_2(t)\frac{\mathrm{d}^2y(t)}{\mathrm{d}t^2} + \cdots + c_R(t)\frac{\mathrm{d}^Ry(t)}{\mathrm{d}t^R} + d(t) = 0$$

Initial Value Constraints (Same as MNN)

$$y(t_0) = u_0, \quad \frac{\mathrm{d}y(t_0)}{\mathrm{d}t} = u_1, \quad \frac{\mathrm{d}^2y(t_0)}{\mathrm{d}t^2} = u_2, \quad \cdots$$

Smoothness Constraints Based on Taylor Expansion

(Different from MNN)

Forward: $y\big(t + s(t)\big) = \sum_{r=i}^{R} \frac{s(t)^r}{r!}\frac{\mathrm{d}^ry(t)}{\mathrm{d}t^r}, \ i = 0, \cdots, R$

Backward: $y(t) = \sum_{r=i}^{R} \frac{\big(-s(t)\big)^r}{r!}\frac{\mathrm{d}^ry\big(t+s(t)\big)}{\mathrm{d}t^r}, \ i = 0, \cdots, R$

S-MNN Solver

Discrete Solutions of $y(t)$ and its Derivatives

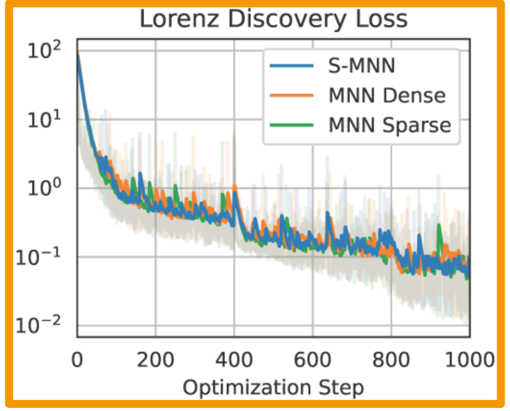
Decoder Network

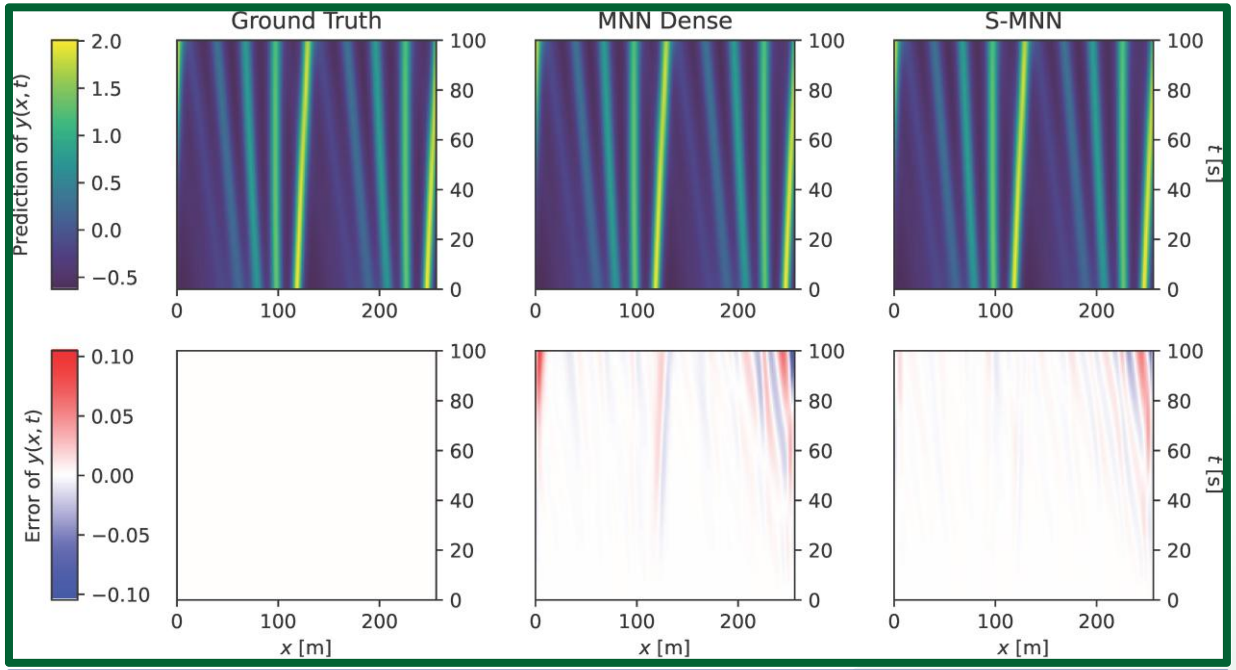
Output Sequence $z_1, z_2, \cdots, z_T$

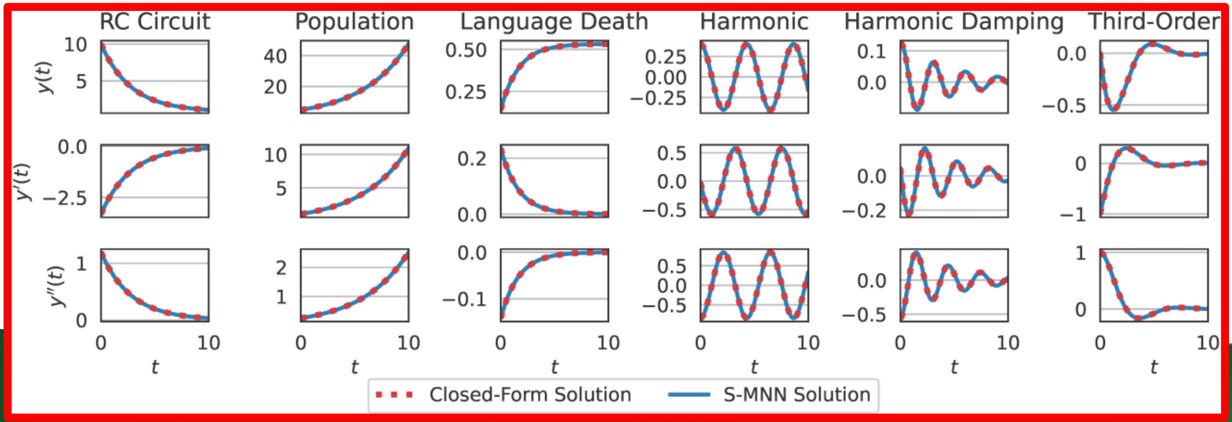
Institute of Science and Technology Austria

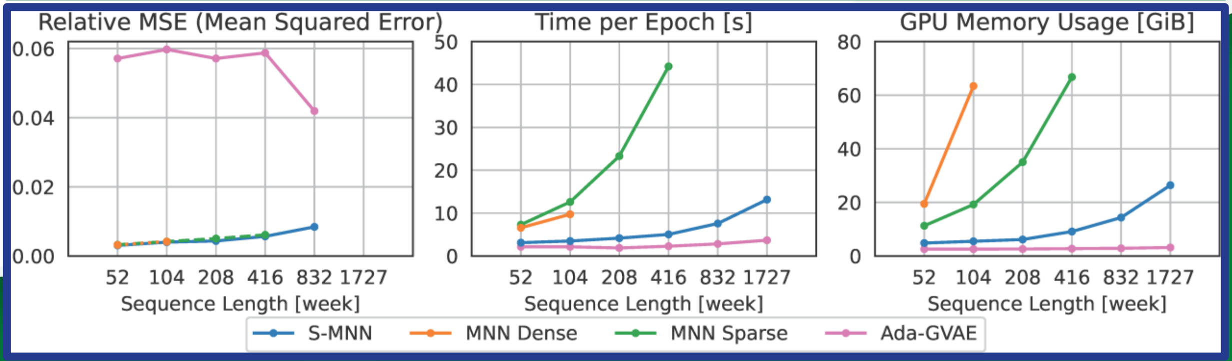
Evaluations

- Standalone validation [ODEs with closed-form solutions]
- Discovery of governing equations [Lorenz system]
- Solving PDEs [Korteweg-De Vries (KdV) equation]
- Real-world application [long-term sea surface temperature forecasting]

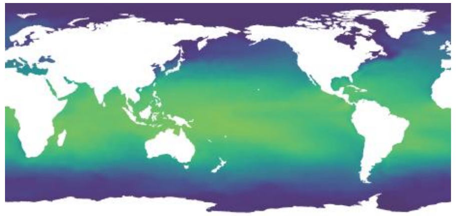








Conclusions



S-MNN overcomes the computational limitations of MNN, enabling scalable, interpretable, and efficient modeling of complex systems.

- Linear time and space complexities with respect to the sequence length
- Applications in chaotic systems, PDEs, and climate modeling
- Advances scientific machine learning

Source code is available at https://github.com/IST-DASLab/ScalableMNN

Institute of Science and Technology Austria