

HeunNet: Extending ResNet using Heun's Method

Mehrdad Maleki
Dept of Computer Science
Maynooth University
Maynooth, Ireland
0000-0003-1020-3322

Mansura Habiba
Dept of Computer Science
Maynooth University
Maynooth, Ireland
0000-0001-9051-1370

Barak A. Pearlmutter
Dept of Computer Science & Hamilton Inst
Maynooth University
Maynooth, Ireland
0000-0003-0521-4553

Abstract—There is an analogy between the ResNet (Residual Network) architecture for deep neural networks and an Euler solver for an ODE. The transformation performed by each layer resembles an Euler step in solving an ODE. We consider the Heun Method, which involves a single predictor-corrector cycle, and complete the analogy, building a predictor-corrector variant of ResNet, which we call a HeunNet. Just as Heun's Method is more accurate than Euler's, experiments show that HeunNet achieves high accuracy with low computational (both training and test) time compared to both vanilla recurrent neural networks and other ResNet variants.

Index Terms—Deep Neural Network, Residual Neural Network, Time Series

I. INTRODUCTION

A ResNet or Residual Network [6] can be viewed as using Euler's Method to integrate a time-dependent ordinary differential equation. Since Euler's Method is very inaccurate, with error = $\mathcal{O}(\Delta t)$, one might conjecture that some of the power of a ResNet with many layers is being devoted to compensating for integration error introduced by this crude approximation. This suggests the use of a more accurate integration method, which might result in less error being introduced, and therefore derive greater benefit from the same number of layers.

Here we use the Heun Method [11], which is much more accurate: error = $\mathcal{O}(\Delta t^2)$. In our HeunNet model, each block computes x_t , the cell state at time t , along with an initial estimate of the next cell state, x_{t+1} . The Heun Method finds x_t using both x_{t+1} and an initial estimate of x_t . This approach helps proposed HeunNet model to provide better result with low memory footprint.

This Heun Method approach can be adapted for a wide variety of layered neural network architectures. HeunNet model only takes input for each state. It then computes both the current and the following cell states using the Heun Method, as shown in Eq. (5).

Each block of the HeunNet neural network uses the same weight and computation method. Therefore, it does not cause additional memory or computation energy. Although each iteration can take longer than simple neural networks, the result shows that the proposed neural network achieves higher accuracy than other neural networks at an early stage of training.

Mehrdad Maleki funded by Irish Research Council (IRC) Fellowship GOIPD/2019/803.

The main objectives of this work are:

- Design a new neural network architecture based on the Heun Method (HeunNet).
- Investigate different evaluation metrics to assess the performance of the proposed model.

II. HEUN'S METHOD FOR SOLVING ODES

Heun's Method is a numerical method for solving ordinary differential equations (ODEs). Heun's Method is sometimes referred to as an "improved Euler" method, but it is better viewed as similar to a two-stage Runge-Kutta method [11]. For solving an ODE $\dot{x} = f(t, x)$ with initial condition $x(0) = x_0$ using Euler's method, we use the slope f at x_0 and move along the tangent line to x at x_0 to reach to estimate x_{t+1} ,

$$x_{t+1} = x_t + hf(t, x_t) \quad (1)$$

where h is the step size. If the actual solution is convex (or concave), the Euler method will underestimate (or overestimate) the next state of the system. So in the long term, the numerical solution diverges from the correct solution. A better estimate can be obtained using an estimate of the slope at the next step and using the average of this slope with the current step's slope. This is the essence of Heun's Method,

$$\tilde{x}_{t+1} = x_t + hf(t, x_t) \quad (2a)$$

$$x_{t+1} = x_t + h \frac{f(t, x_t) + f(t+1, \tilde{x}_{t+1})}{2} \quad (2b)$$

In predictor-corrector methods like this there are two steps, "prediction" and "correction." A "predictor" (2a) estimates the value at the next step, and this estimate is improved by a "corrector" (2b). The Euler and Heun Method solutions for the ODE $\dot{x} = 2\sqrt{x}$ with initial condition $x(0) = 1$ are shown in Fig. 1.

III. MOTIVATION

Consider a neural network with L layers such that the number of neurons in all the layers are equal. Let X_0 be the input of the network and let W_0 be the square matrix (since all layers have the same number of neurons) of weight between the input and the first layers. Let $Z_1 = W_0 X_0$ and $X_1 = \sigma(Z_1)$. So we define $\mathcal{F}(X_k) = \sigma(W_k X_k) = X_{k+1}$. The regular architecture of a neural network is straightforward as shown in Eq. (3):

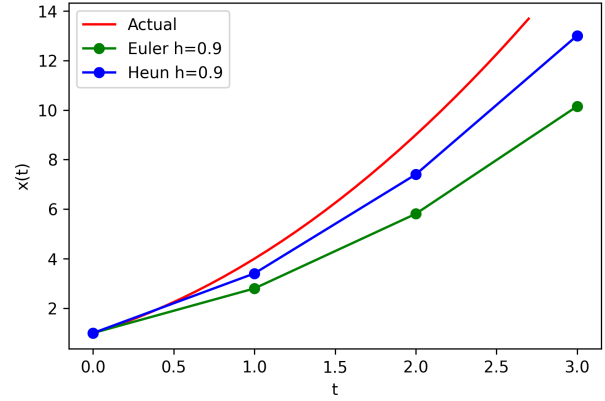
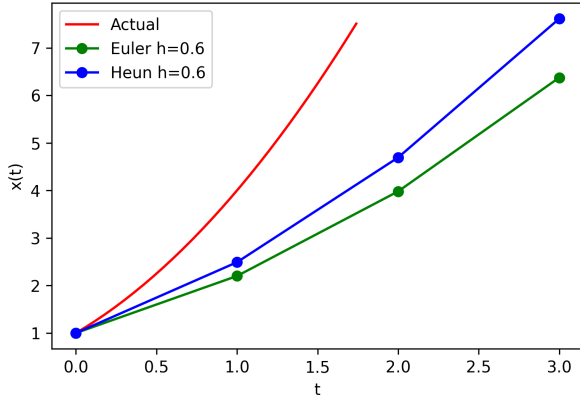


Fig. 1: Euler vs Heun Method for $\dot{x} = 2\sqrt{x}$ with $x(0) = 1$, and $h = 0.6$ (left) or $h = 0.9$ (right). The analytic solution of this ODE is $x(t) = (t + 1)^2$ (red curve).

$$X_{k+1} = \mathcal{F}(X_k) \quad (3)$$

A residual neural network (ResNet) uses skip connections to deal with the problem of vanishing gradients during the training.

$$X_{k+1} = X_k + \mathcal{F}(X_k) \quad (4)$$

This architecture is quite efficient when dealing with an intense neural network, even in combination with convolutional layers. However, if we look closely, we realize that this is similar to Euler's method for solving an ODE with step size $h = 1$ [2]. The motivation is that ResNet is a modification of the Euler method, so HeunNet is the modification of Heun's Method.

IV. PROPOSED MODEL: HEUNNET MODEL

Our motivation inspired by the Heun Method [11]. Since ResNet [6] is a discretization of the Euler method, we could see the Heun Method as the extension of ResNet. In a nutshell,

$$\boxed{\text{ResNet : Euler's Method} :: \text{HeunNet : Heun's Method}}$$

We propose another architecture for neural networks motivated by the Heun Method for solving ODE. In our model, the output of the next layer is a combination of the output from ResNet and the output of the previous layer as shown in Eq. (5a).

$$\tilde{X}_{k+1} = X_k + \mathcal{F}(X_k) \quad (5a)$$

$$X_{k+1} = X_k + \frac{1}{2}(\mathcal{F}(X_k) + \mathcal{F}(\tilde{X}_{k+1})) \quad (5b)$$

Fig. 2 shows the essence of this architecture.

Residual Neural Network is a discretization of the Euler method. In the residual network, the state of the layer $k + 1$ is related to the state of the layer k as shown in (4), where \mathcal{F} can be an arbitrary layer transition function. But if we use Heun's Method for solving $\dot{x}(t) = \mathcal{F}(x(t))$ with initial condition $x(0) = x_0$ and let step size be $h = 1$, we obtain the architecture of the Heun Neural Network (HeunNet), i.e., Eq. (5).

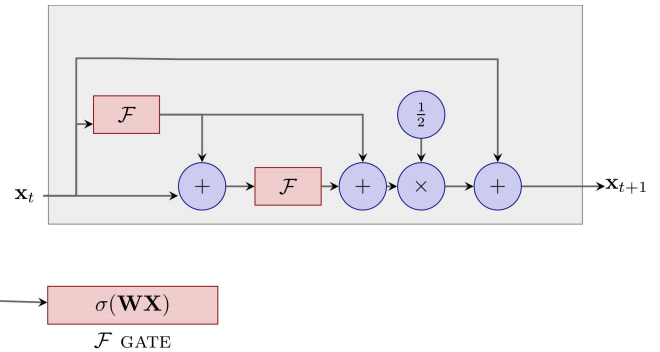


Fig. 2: The architecture of proposed Heun based Recurrent Neural Network (HeunNet)

The Heun Method has quadratic accuracy, while the accuracy of Euler's method is linear. For this reason, the Heun Method is generally preferred to Euler's method when numerically solving ODEs. A similar situation is occurring for the Heun Neural Network (HeunNet), which appears to converge to a more accurate result in fewer layers.

A. Gradient Propagation in a HeunNet

The original motivation for ResNet was to avoid the vanishing gradient problem by causing the Jacobian of the transition between layer k , and layer $k + 1$ to be a near-identity matrix. Both the ResNet equation (4) and the analogous HeunNet equation (5) are of the form $X_{k+1} = X_k + \text{something}$, which accomplishes this. Like ResNet, HeunNet is an entirely feedforward computational process that can be expressed using standard numeric linear algebra routines. As such, it is straightforward to implement in any deep learning framework, such as PyTorch. The gradient can be automatically calculated via backpropagation, i.e., reverse mode Automatic Differentiation [1], and that process is just as automatic as for ResNet.

Despite its conceptual similarity, and as we see below, numerical simulations show that HeunNet obtains better performance than ResNet across various tasks.

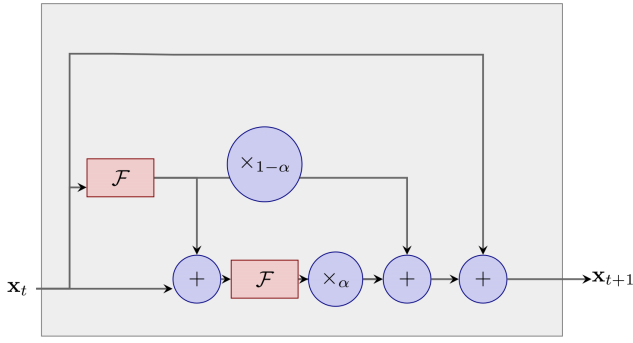


Fig. 3: The architecture of proposed Extended Heun based Recurrent Neural Network (ExtendedHeunNet)

V. EXTENSION OF HEUNNET MODEL

In the Heun's Method for solving ODEs, if we put more weights on the corrector's second term we will have a better approximation of the actual function. In the Heun's Method, we first compute the $f(x_k)$ as the slope of the current state and $f(\tilde{x}_{k+1})$ as the approximation coming from Euler method and get the average of these two as the corrector. But if we put more weights on the $f(\tilde{x}_{k+1})$ and use convex hull of $f(x_k)$ and $f(\tilde{x}_{k+1})$, i.e., $(1-\alpha)f(x_k) + \alpha f(\tilde{x}_{k+1})$ with α near to 1, then we have better approximation than Heun's Method:

$$\tilde{x}_{k+1} = x_k + h f(x_k) \quad (6a)$$

$$x_{k+1} = x_k + h ((1-\alpha)f(x_k) + \alpha f(\tilde{x}_{k+1})) \quad (6b)$$

where $0 \leq \alpha \leq 1$ and h is the step size. Fig. 1 shows the difference between Heun's Method and extended Heun's Method for $\dot{x} = 2\sqrt{x}$ with $x(0) = 1$.

For $h = 1$ we obtain the architecture of a Neural Network which is an extension of HeunNet (ExtendedHeunNet), i.e.,

$$\tilde{x}_{k+1} = x_k + f(x_k) \quad (7a)$$

$$x_{k+1} = x_k + (1-\alpha)f(x_k) + \alpha f(\tilde{x}_{k+1}) \quad (7b)$$

- If $\alpha = 0$ then $x_{k+1} = \tilde{x}_{k+1}$ and we have **ResNet**.
- If $\alpha = \frac{1}{2}$ then we have **HeunNet**.
- If $\alpha \neq \frac{1}{2}$ then we have **ExtendedHeunNet**.

The question is, which values of α gives us better approximation? Does this modification improves the performance of HeunNet? Since we need more weights on the next state to have a better approximation of the actual function, we choose α to be close to 1. The architecture of this extended Neural Network is shown in Fig. 3.

VI. RESULTS

For performance evaluation of proposed new model, we have chosen two tasks: (i) MNIST dataset classification [4], and (ii) ECG classification [8]. We compared the HeunNet model with other Recurrent Neural Network, e.g., LSTM [7] and GRU [3].

TABLE I: Parameters of Dataset Used

#	Parameter	Value
1	Number of Samples	109446
2	Number of Categories	5
3	Sampling Frequency	125 Hz
4	Classes	[N: 0, S: 1, V: 2, F: 3, Q: 4]

A. MNIST classification Task

Fig. 4 shows the comparative analysis of loss and accuracy for LSTM and HeunNet models over ten iterations. LSTM achieves an accuracy of 95.44% after the 10th iteration. On the other hand, the HeunNet model achieves an accuracy of 98.26% at the 10th iteration.

B. ECG heartbeat classification Task

For this task, we used the dataset for MIT-BIH Arrhythmia Database v1.0.0 PhysioNet [5]. Table I shows some characteristics of the used datasets. This dataset consists of signals that correspond to electrocardiogram (ECG) shapes of heartbeats for the normal case and shapes caused by arrhythmias and other myocardial infarctions. There are five categories of beats present in the dataset as Nonectopic beat (N), Supraventricular ectopic beat (S), Ventricular ectopic beat (V), Fusion beat (F) and Unknown beat (Q).

Fig. 5 shows the loss and accuracy for LSTM model over 100 iterations. The best iteration for the LSTM model in the ECG heartbeat classification task is 79 with an accuracy 90.40%.

With the HeunNet model, we get the best accuracy of 98.80%. Fig. 5 shows the loss and accuracy for LSTM model for only 50 iterations.

The proposed model can achieve higher accuracy within only half of the number of iterations required for other recurrent neural networks.

C. Time Series prediction

As shown in (5), the neural network \mathcal{F} can be any neural network. For time series generation task, we used a simple LSTM neural network for \mathcal{F} in (5). This Heun neural network-based LSTM model is compared against Phased LSTM [9] and vanilla LSTM model. A sine wave of length 16π is used for the dataset for this task. Each model is trained for 100 iterations. Fig. 7 shows the comparative performance analysis for three models. As shown in Fig. 7, the HeunNet model outperformed the other two model.

Proposed ExtendedHeunNet model's prediction is more accurate than other HeunNet and also other two neural network models. ExtendedHeunNet provides an accuracy of 97.30% with $\alpha = 0.8$ in just 2 iterations for Task C. For ExtendedHeunNet, $0.75 < \alpha \leq 0.8$ provides best accuracy. If α move towards 1, the accuracy starts to drop.

Fig. 8 shows generated sine wave for LSTM, Phased-LSTM and HeunNet model. The HeunNet and ExtendedHeunNet model's prediction are more accurate than the predictions of the other two models.

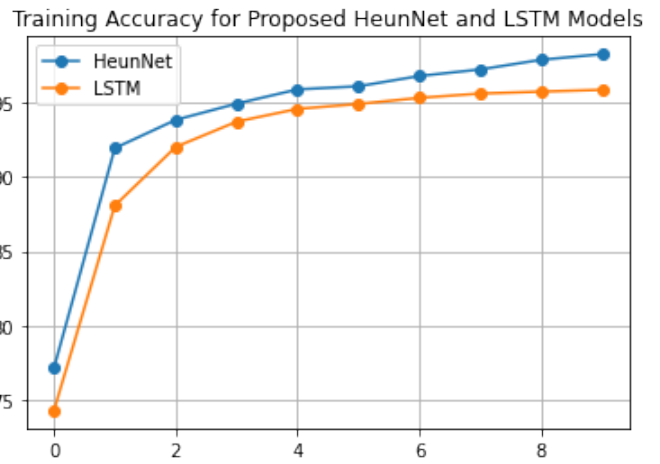
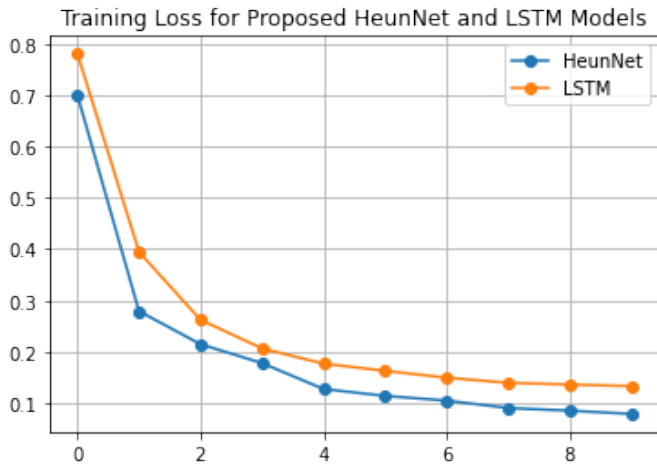


Fig. 4: HeunNet and LSTM model loss (left) and accuracy (right) for 10 iterations on MNIST classification task

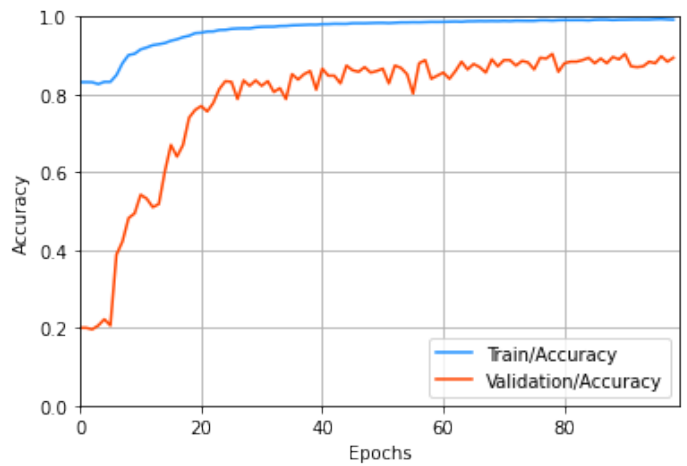
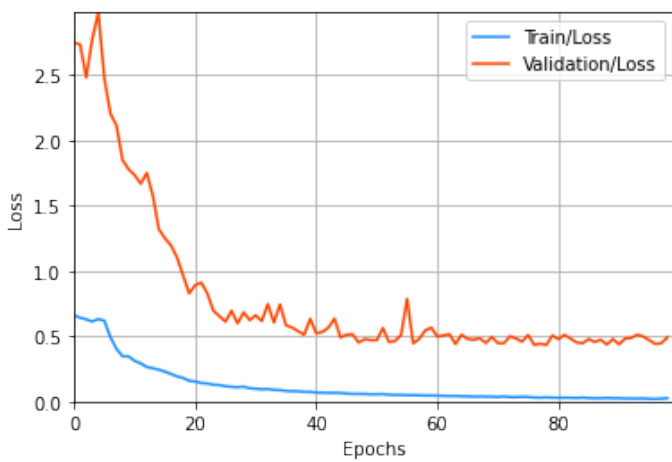


Fig. 5: LSTM model Loss (left) and accuracy (right) for 100 iterations on ECG heartbeat classification task

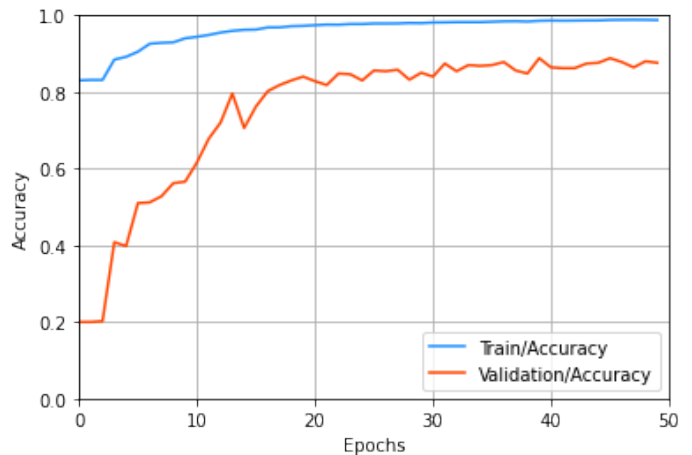
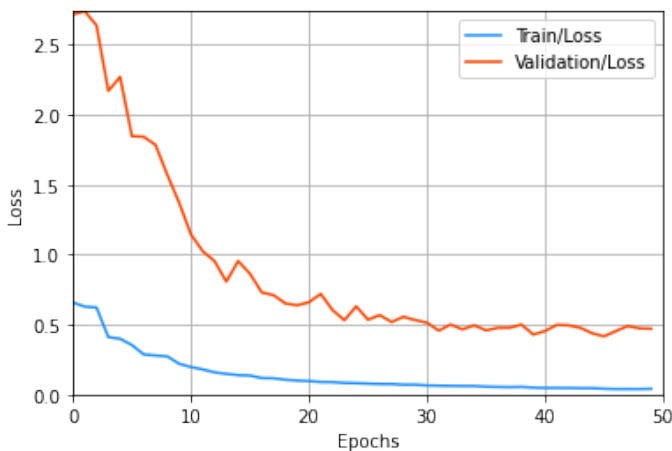


Fig. 6: HeunNet model loss (left) and accuracy (right) for 50 iterations on ECG heartbeat classification task

D. Performance Evaluation

Typically, the accuracy of RNN models (LSTM or GRU) increase with the increase in the number of iteration. There-

fore, the training time is comparatively long. We experimented with different ranges of iteration for different problems. For task A, within ten iterations, LSTM achieve 95.44% accuracy,

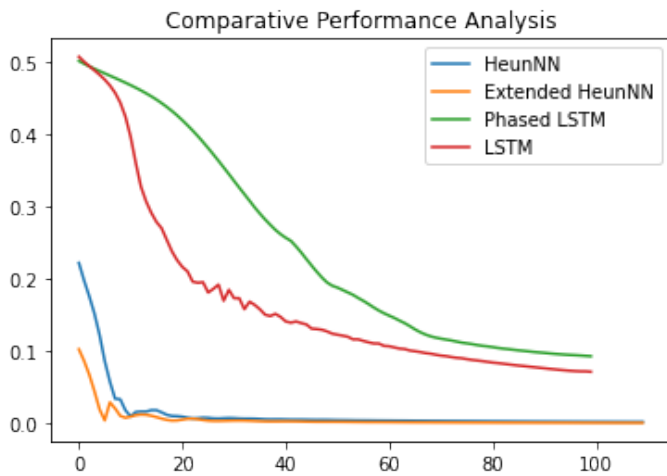


Fig. 7: Loss for LSTM, Phased-LSTM, and HeunNet models over 100 iterations for Sine Wave generation

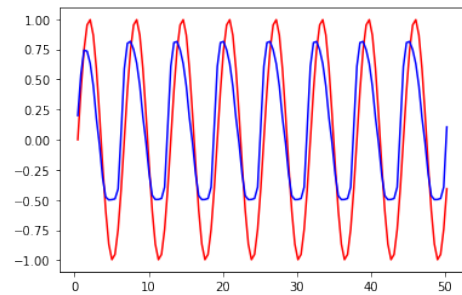
and GRU achieves 92.01% accuracy. However, for the same number of iteration, the HeunNet model achieves 97.98% accuracy. The result shows that the proposed model needs less training time to achieve high accuracy.

In a traditional bi-directional vanilla recurrent neural network, corresponding parameters are combined into a single LSTM. On the other hand, in this proposed model, the results of two individual neural networks are combined using the Heun Method (5).

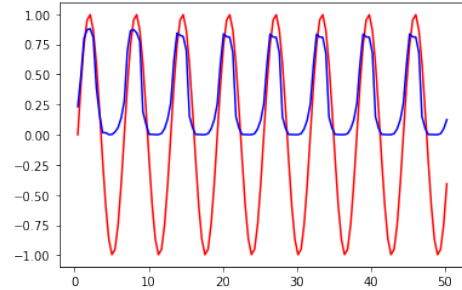
In this section, both classification and time series prediction tasks are evaluated for different recurrent models against the HeunNet model. We can conclude that the proposed model can achieve higher accuracy with almost half the number of iterations than other models. The HeunNet model does not need to train for a long time to get the optimized result.

VII. DISCUSSION

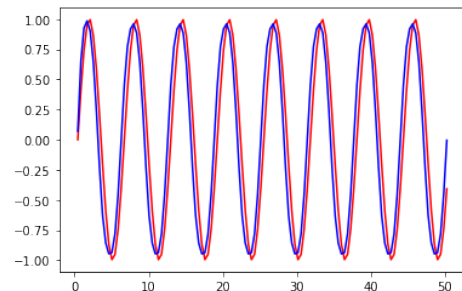
ResNet has the form of an Euler method for discretizing and solving ODEs, and HeunNet and its extensions can be similarly viewed as having the form of Heun's Method. Some other groups have independently explored similar ideas for extending ResNet based on other numeric ODE solvers. In particular, MomentumNets [10] are constructed using an analogy with a numeric solver for second-order ODEs, while our HeunNet and its extension are derived from solution methods for first-order ODEs. Since second-order ODEs can be reduced to first-order ODEs using standard techniques, it should be possible to put these all into a common framework. MomentumNets have ResNet as a special case, but our extended HeunNet has both ResNet and HeunNet as special cases. And finally, in MomentumNets the velocity term is defined as a function of the current state in the first stage, and in the second stage this term is combined with the input of the previous layer, while our HeunNet or its extension use ResNet as the approximation for the next state and the second term is a correction for the error introduced in the first step. So the HeunNet first makes an approximation for the next state, and a weighted average



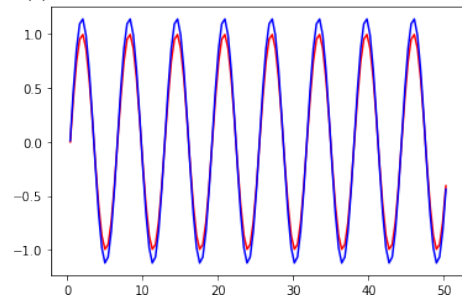
(a) LSTM model trained for 100 iterations



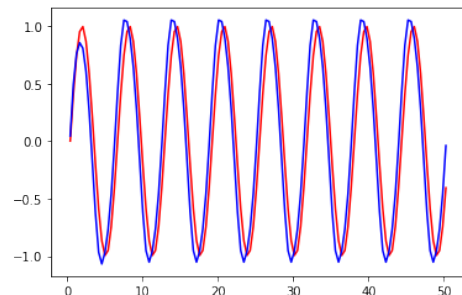
(b) Phased-LSTM model trained for 100 iterations



(c) HeunNet model trained for 25 iterations



(d) ExtendedHeunNet model trained for 2 iterations with $\alpha = 0.8$



(e) ExtendedHeunNet model trained for 10 iterations with $\alpha = 0.8$

Fig. 8: Training different models to generate a sine wave.

of this approximation and its correction constitutes the space of candidates for the next state of the system.

Heun's Method falls in the general class of predictor-corrector methods: it makes an initial guess as to the correct state of the next time step, then corrects that estimate using more accurate information based on this first estimate. This can be viewed as a local causality violation, as it makes the update between layers t and $t + 1$ depend not just on the state of layer t , but also on that of layer $t + 1$. However the method remains causal, unlike time-series processing techniques like a Bi-directional RNN, which consists of a forward-in-time pass along with a corresponding backward-in-time pass. The technique proposed here might be naturally extended to that setting by making each of these passes update using a Heun's Method step rather than an Euler step. So we should not view these as in competition, but rather as potentially synergistic.

VIII. CONCLUSION

We propose a new neural network model that leverages Heun's Method for hidden state optimization. We also developed a general architecture that includes the HeunNet as a special case. Performance evaluation demonstrates that both proposed models outperform existing recurrent neural networks on a few problems. The proposed models also take less time for training, and achieve higher accuracy. These architectures leverage the strength of Heun's Method, which provides higher accuracy. The error generated by the first step of the Heun Method is compensated by the corrector, i.e., the second term, thus reducing the error [11].

REFERENCES

- [1] Atılım Güneş Baydin, Barak A. Pearlmutter, Alexey Andreyevich Radul, and Jeffrey Mark Siskind. Automatic differentiation in machine learning: a survey. *Journal of machine learning research*, 18, 2018.
- [2] Ricky T. Q. Chen, Yulia Rubanova, Jesse Bettencourt, and David Duvenaud. Neural ordinary differential equations. *arXiv preprint arXiv:1806.07366*, 2018.
- [3] Kyunghyun Cho, Bart Van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. Learning phrase representations using RNN encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078*, 2014.
- [4] Li Deng. The MNIST database of handwritten digit images for machine learning research [best of the web]. *IEEE Signal Processing Magazine*, 29(6):141–142, 2012.
- [5] Ary L. Goldberger, Luis A. N. Amaral, Leon Glass, Jeffrey M. Hausdorff, Plamen Ch. Ivanov, Roger G. Mark, Joseph E. Mietus, George B. Moody, Chung-Kang Peng, and H. Eugene Stanley. Physiobank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. *Circulation*, 101(23):e215–e220, 2000.
- [6] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.
- [7] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.
- [8] Loren Kersey, Kat Lilly, and Noel Park. ECG heartbeat classification: An exploratory study. In *Proceedings of the Australasian Joint Conference on Artificial Intelligence-Workshops*, pages 20–27, 2018.
- [9] Daniel Neil, Michael Pfeiffer, and Shih-Chii Liu. Phased LSTM: Accelerating recurrent network training for long or event-based sequences. *arXiv preprint arXiv:1610.09513*, 2016.
- [10] Michael E Sander, Pierre Ablin, Mathieu Blondel, and Gabriel Peyré. Momentum residual neural networks. *arXiv preprint arXiv:2102.07870*, 2021.
- [11] Dennis G Zill. *A first course in differential equations with modeling applications*. Cengage Learning, 2012.