ESANN 2023 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 4-6 October 2023, i6doc.com publ., ISBN 978-2-87587-088-9. Available from http://www.i6doc.com/en/.

Residual Reservoir Computing Neural Networks for Time-series Classification

Andrea Ceni and Claudio Gallicchio *

Department of Computer Science, University of Pisa
Largo Bruno Pontecorvo 3 - 56127 Pisa, Italy

Abstract. We introduce a novel class of Reservoir Computing (RC) models, a family of efficiently trainable Recurrent Neural Networks based on untrained connections. Aiming to improve the forward propagation of input information through time, we augment standard Echo State Networks (ESNs) with linear reservoir-skip connections modulated by an untrained orthogonal weight matrix. We analyze the mathematical properties of the resulting reservoir systems and show that the dynamical regime of the proposed class of models is controllably close to the edge of stability. Experiments on several time-series classification tasks highlight the striking performance advantage of the proposed approach over standard ESNs.

1 Introduction

Reservoir Computing (RC) networks [1] are a class of recurrent neural models that have become extremely popular over the years due to their efficient training. Rather than applying end-to-end backpropagation through time training, RC exploits the properties of asymptotically stable recurrent layers to avoid the computational burden of training algorithms as much as possible. In fact, the only trainable component in the architecture is a readout layer. In practice, the hidden recurrent layer of the architecture, called *reservoir*, remains untrained after random initialization, subject to a stability condition known as the echo state property [2]. On the one hand, this property allows for stable dynamics and well-behaved state-space organization, which has been successfully exploited in various application scenarios, especially in pervasive AI environments. On the other hand, the system is intrinsically biased towards fading memory computation, which inevitably reduces the ability of the network to effectively propagate the driving input information across multiple time steps.

In this paper, we address the problem of effective information propagation in untrained dynamical neural systems. We do so by introducing a new class of RC models that essentially modify the reservoir architecture by introducing skip connections that linearly propagate the network state to the next time step (an idea also supported by biological plausibility [3]). From an architectural point of view, these connections are introduced in the same spirit as those found in Residual Neural Networks [4]. However, in our case, the introduction is concerned with the nature of the temporal propagation of the reservoir state rather than the residual nature of the learning problem. Therefore, in order to maximize the information content of the time-propagated state memory, the residual

*This work is partially supported by the EC H2020 programme under project TEACHING (grant n. 871385), and by EMERGE, a project funded by EU Horizon research and innovation programme (grant n. 101070918).

ESANN 2023 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 4-6 October 2023, i6doc.com publ., ISBN 978-2-87587-088-9. Available from http://www.i6doc.com/en/.

reservoir connections are modulated by an untrained orthogonal weight matrix, thus exploiting, in an RC context, the optimal memory properties of this type of dynamic neural systems [5].

We investigate variants of the proposed residual RC approach, analyze their stability properties, and provide experimental evidence for the substantial performance advantage that can be achieved in time-series classification tasks over standard RC methodology.

2 Reservoir Computing

As a fundamental baseline for our proposal, we introduce the Echo State Network (ESN) model [6], which in its vanilla formulation includes a fixed non-linear reservoir layer and a trainable readout. We consider, in particular, the general case with leaky-integrator recurrent neurons [7]. The state transition equation of the reservoir is given as follows:

$$\mathbf{h}(t) = (1-\alpha)\ \mathbf{h}(t-1) + \alpha\ \phi(\mathbf{W_h}\mathbf{h}(t-1) + \mathbf{W_x}\mathbf{x}(t) + \mathbf{b}), \tag{1}$$

where $\mathbf{h}(t) \in \mathbb{R}^{N_h}$ and $\mathbf{x}(t) \in \mathbb{R}^{N_x}$ respectively denote the state and input at time t, $\mathbf{W_h}$ and $\mathbf{W_x}$ are the reservoir and the input weight matrices, $\mathbf{b}$ is the bias vector, $\phi(\cdot)$ is an element-wise applied non-linearity (we use $\tanh(\cdot)$), and $\alpha \in (0, 1]$ is a leakage hyper-parameter. The reservoir is typically initialized in the origin, i.e., $\mathbf{h}(0) = \mathbf{0}$. The values in $\mathbf{W_h}$ are randomly chosen and then rescaled to have a specific value of the spectral radius (i.e., the maximum length of an eigenvalue), a crucial hyper-parameter denoted as ρ. $\mathbf{W_x}$ and $\mathbf{b}$ are randomly initialized from uniform distributions over $(-\omega_x, \omega_x)$, and $(-\omega_b, \omega_b)$, where ω_x and ω_b act respectively as input and bias scaling hyper-parameters. The value of ρ is important as it practically determines the dynamic regime of the reservoir layer, and in applications it is often controlled to values not exceeding 1.

The ESN architecture also comprises a tunable readout layer, which is typically linear and trained in closed-form by ridge regression. For time-series classification problems, the reservoir is run on each input sequence and the state calculated for the last time-step is used to feed the readout classifier.

3 Residual Echo State Networks

We introduce a class of RC models based on reservoirs with linear skip connections in the state processing, and called *Residual Echo State Network* (ResESN). The state transition function of the residual reservoir system is given as follows:

$$\mathbf{h}(t) = \alpha\, \mathbf{O}\ \mathbf{h}(t-1) + \phi(\mathbf{W_h}\mathbf{h}(t-1) + \mathbf{W_x}\mathbf{x}(t) + \mathbf{b}), \tag{2}$$

where $\mathbf{O}$ is a randomly generated orthogonal matrix, and α is a scaling coefficient that we treat as hyper-parameter. As we will show later, the value of α can be used to adjust the dynamic behavior of the reservoir. All other terms in eq. (2) are the same as in eq. (1). With $\alpha = 0$ a traditional ESN [6] is obtained.

Intuitively, considering the application of the recurrent layer over time, the introduction of the skip connections in the additive part of eq. (2) (first term on

ESANN 2023 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 4-6 October 2023, i6doc.com publ., ISBN 978-2-87587-088-9. Available from http://www.i6doc.com/en/.

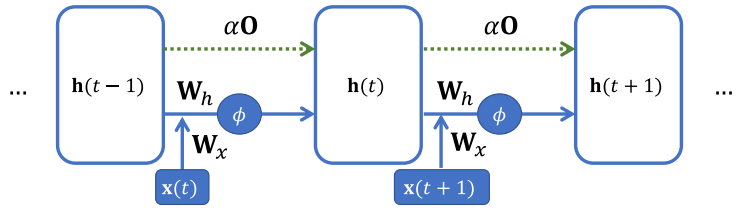

Fig. 1: Computation performed by the reservoir of a ResESN unfolded over time.

the right-hand side) allows the creation of a path for the long-term propagation of the input information. This concept is illustrated graphically in Fig. 1. As typical for the stability analysis of reservoir systems, we consider bias-free autonomous dynamics linearized around the origin, and study the eigenspectrum of the resulting Jacobian. The resulting Jacobian is given by:

$$\mathbf{J} = \alpha\mathbf{O} + \mathbf{W_h}, \tag{3}$$

from which we can make interesting observations. First, we can notice that the eigenvalues of the $\alpha\mathbf{O}$ term are distributed in the complex plane over a circle of radius α centered at the origin. Then, by virtue of the Bauer-Fike's theorem [8], we can identify the position of the eigenvalues of the complete Jacobian $\mathbf{J}$ as being at maximum distance $\|\mathbf{W_h}\|$ from those of $\alpha\mathbf{O}$, i.e., within a $\|\mathbf{W_h}\|$-tube around the circle of radius α. It is interesting to note that this characterization results in an architectural bias of the residual reservoir toward dynamics intrinsically close to the edge of stability [9], represented by eigenvalues on the circle of radius one. The hyper-parameter α, in this context, makes it possible to decide how close to be to this boundary, with a value of $\alpha = 1$ yielding the boundary condition. At the same time, the magnitude of the weights in $\mathbf{W_h}$ determines perturbations to the eigenvalues of $\alpha\mathbf{O}$, allowing a wider range of dynamics to be covered.

In the following, we explore two variants to the residual reservoir introduced in this paper. The first implements as the orthogonal matrix $\mathbf{O}$ of eq. (2) a permutation matrix that implements a circular shift, i.e., a matrix that has ones in the main sub-diagonal and in the upper right-hand corner, and in which all other elements are zero. This variant is denoted ResESN$_C$, and it allows us to simplify the construction of the orthogonal matrix, making it deterministic (with a fixed topology and a single weight value on all non-zero connections) and close to the nature of possible neuromorphic and embedded implementations. The second variant we consider is to replace the orthogonal matrix $\mathbf{O}$ in eq. 2 with the identity matrix $\mathbf{I}$. In this case, the state transition function reads as $\mathbf{h}(t) = \alpha\mathbf{h}(t-1) + \phi(\mathbf{W_h}\mathbf{h}(t-1) + \mathbf{W_x}\mathbf{x}(t) + \mathbf{b})$, which thus has a very similar form to eq. (1), although the linear and non-linear components do not appear as a convex combination (as they do in eq. (1)). We will refer to this second variant hereafter as ResESN$_I$. Note that while the Jacobian eigenspectrums of a ResESN with a random $\mathbf{O}$ orthogonal matrix and a ResESN$_C$ share a similar structure, the case of a ResESN$_I$ is different. In this case, the Jacobian has the form $\alpha\mathbf{I} + \mathbf{W_h}$, whose eigenvalues all lie in a $\|\mathbf{W_h}\|$-neighborhood of α.

ESANN 2023 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 4-6 October 2023, i6doc.com publ., ISBN 978-2-87587-088-9. Available from http://www.i6doc.com/en/.

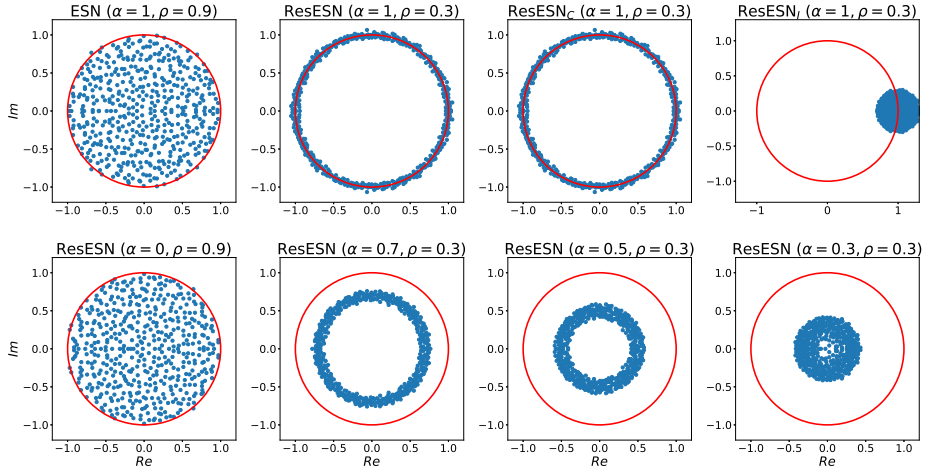

Fig. 2: Eigenvalues of the resulting Jacobian in (autonomous) reservoirs with $N_h = 500$ recurrent neurons. **Top:** Comparison of ESN and variants of ResESN. **Bottom:** Examples of configurations in ResESN by varying hyper-parameters.

An illustration of the positioning of the ResESN eigenvalues is given in Fig. 2. The first row (top) compares configurations obtainable with ESN, ResESN, ResESN$_C$, and ResESN$_I$. Note in particular how ResESN$_C$, although characterized by simplified residual connections, has a similar eigenvalue placement as ResESN in the same configuration. The second row (bottom) shows some examples of configurations that can be obtained by varying the hyperparameters of a ResESN. Note, among other things, how it is possible to reproduce the configuration of ESN (first column, top row in the figure), and how it is possible to model different levels of proximity to the circle of radius one.

4 Experiments

We experimentally validate the proposed methodology on a set of diverse time-series classification benchmarks from the UEA & UCR repository[1]. We used the original split in training and test sets, applying a further 67%-33% stratified splitting of the whole training data into training and validation sets. Table 1 provides a summary of the main properties of the considered datasets.

We ran experiments with ResESN and its introduced variants, considering a number of reservoir neurons N_h in the range 10-500. We explored values of α and ρ in $[0, 1]$, values of ω_x and ω_b in $\{10^{-3}, 10^{-2}, \ldots, 10\}$. For $\mathbf{W_h}$, we have used the fast initialization method introduced in [10]. For comparison, we also ran experiments with the vanilla ESN baseline, with the same range of configurations. Individually for each model, the hyper-parameters were fine-tuned by model selection, using random search with 500 iterations. After model

[1] www.timeseriesclassification.com

ESANN 2023 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 4-6 October 2023, i6doc.com publ., ISBN 978-2-87587-088-9. Available from http://www.i6doc.com/en/.

selection, the selected configuration was trained on the whole training set, and assessed on the test set, considering 10 random guesses. The results reported in the following are averaged over these random guesses. The code was written in Keras and scikit-learn[2], and was run on a system with 2x20 Intel(R) Xeon(R) CPU E5-2698 v4 @ 2.20GHz.

Name	**# Tr**	**# Ts**	**Length**	**Input dim**	**# Classes**
Adiac	390	391	176	1	37
Blink	500	450	510	4	2
CinCECGTorso	40	1380	1639	1	4
FordA	3601	1320	500	1	2
FordB	3636	810	500	1	2
KeplerLightCurves	920	399	4767	1	7
Libras	180	180	45	2	15
Lightning2	60	61	637	1	2
OliveOil	30	30	570	1	4
ShapesAll	600	600	512	1	60
StarLightCurves	1000	8236	1024	1	3
UWaveGestureLibraryAll	896	3582	945	1	8
Wafer	1000	6164	152	1	2
Yoga	300	3000	426	1	2

Table 1: Summary of the datasets used.

The achieved results are reported in Table 2. The values shown in the table clearly highlight the striking advantage of the class of models proposed in this paper over the traditional ESN approach, surpassing its accuracy in all the cases analyzed, sometimes by a considerable margin. In particular, we can observe that the base ResESN achieves the best result in the vast majority of cases. Moreover, despite architectural simplifications, ResESN$_C$ generally achieves an accuracy close to that of ResESN. Among the variants analyzed, although ResESN$_I$ is the one that generally leads to the worst results, it also succeeds in improving (in some cases significantly) the performance of ESN. This last observation further attests to the effective flexibility provided by the absence of convex combinations in the state transition equation of ResESN.

5 Conclusions

We have introduced Residual Echo State Networks (ResESNs), a novel class of randomized recurrent neural networks, under the umbrella of Reservoir Computing. Our proposal extends the vanilla Echo State Network (ESN) architecture, introducing linear reservoir-skip connections modulated by a random orthogonal weight matrix. The idea behind our proposal is that these orthogonal residual connections enable the effective propagation of input information across multiple time-steps, making the approach particularly suitable for time-series classification problems. Our mathematical analysis showed the flexibility of the method in generating dynamics that are controllably close to the border of stability.

[2]Each RC architecture is implemented as a custom Keras model, where the readout is a scikit-learn ridge classifier.

ESANN 2023 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 4-6 October 2023, i6doc.com publ., ISBN 978-2-87587-088-9. Available from http://www.i6doc.com/en/.

Dataset	**ESN**	**ResESN**	**ResESN$_C$**	**ResESN$_I$**
Adiac	$0.278_{\pm 0.004}$	$0.554_{\pm 0.028}$	$0.550_{\pm 0.029}$	$\mathbf{0.609}_{\pm 0.013}$
Blink	$0.622_{\pm 0.014}$	$0.648_{\pm 0.038}$	$0.639_{\pm 0.030}$	$\mathbf{0.764}_{\pm 0.033}$
CinCECGTorso	$0.260_{\pm 0.001}$	$\mathbf{0.427}_{\pm 0.065}$	$0.294_{\pm 0.021}$	$0.253_{\pm 0.010}$
FordA	$0.534_{\pm 0.012}$	$0.691_{\pm 0.015}$	$\mathbf{0.713}_{\pm 0.037}$	$0.596_{\pm 0.041}$
FordB	$0.519_{\pm 0.002}$	$0.542_{\pm 0.037}$	$\mathbf{0.564}_{\pm 0.014}$	$0.538_{\pm 0.005}$
KeplerLightCurves	$0.354_{\pm 0.025}$	$\mathbf{0.561}_{\pm 0.018}$	$0.522_{\pm 0.045}$	$0.420_{\pm 0.036}$
Libras	$0.493_{\pm 0.018}$	$0.769_{\pm 0.019}$	$\mathbf{0.801}_{\pm 0.011}$	$0.698_{\pm 0.024}$
Lightning2	$0.610_{\pm 0.007}$	$\mathbf{0.644}_{\pm 0.018}$	$0.623_{\pm 0.007}$	$0.611_{\pm 0.010}$
OliveOil	$0.400_{\pm 0.000}$	$0.813_{\pm 0.054}$	$\mathbf{0.847}_{\pm 0.045}$	$0.633_{\pm 0.058}$
ShapesAll	$0.499_{\pm 0.006}$	$\mathbf{0.706}_{\pm 0.009}$	$0.705_{\pm 0.010}$	$0.590_{\pm 0.017}$
StarLightCurves	$0.867_{\pm 0.001}$	$\mathbf{0.931}_{\pm 0.004}$	$0.929_{\pm 0.002}$	$0.879_{\pm 0.009}$
UWaveGestureLibraryAll	$0.757_{\pm 0.006}$	$\mathbf{0.891}_{\pm 0.014}$	$0.874_{\pm 0.012}$	$0.850_{\pm 0.013}$
Wafer	$0.977_{\pm 0.001}$	$\mathbf{0.990}_{\pm 0.002}$	$0.988_{\pm 0.003}$	$0.981_{\pm 0.003}$
Yoga	$0.619_{\pm 0.005}$	$0.725_{\pm 0.015}$	$0.742_{\pm 0.019}$	$\mathbf{0.754}_{\pm 0.006}$

Table 2: Test set accuracy. Best results for each dataset are highlighted in bold.

Experiments confirmed the goodness of the ResESN approach, showing its surprising effectiveness compared to traditional ESNs. Moreover, the analysis of architectural variants highlighted the good performance obtained by residual connections corresponding to a simple matrix implementing a circular shift.

Beyond the already excellent results obtained in this work, we believe that future research deserves to further explore architectures based on residual reservoirs, also studying their (possibly local) adaptation algorithms and extensions to larger domains such as graphs and temporal graphs.

References

[1] K. Nakajima and I. Fischer. *Reservoir Computing*. Springer, 2021.

[2] I. B. Yildiz, H. Jaeger, and S. J Kiebel. Re-visiting the echo state property. *Neural networks*, 35:1–9, 2012.

[3] Q. Liao and T. Poggio. Bridging the gaps between residual learning, recurrent neural networks and visual cortex. *arXiv preprint arXiv:1604.03640*, 2016.

[4] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[5] O. L. White, D. D. Lee, and H. Sompolinsky. Short-term memory in orthogonal neural networks. *Physical review letters*, 92(14):148102, 2004.

[6] H. Jaeger and H. Haas. Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication. *science*, 2004.

[7] H. Jaeger, M. Lukoševičius, D. Popovici, and U. Siewert. Optimization and applications of echo state networks with leaky-integrator neurons. *Neural networks*, 20(3):335–352, 2007.

[8] F. L. Bauer and C. T. Fike. Norms and exclusion theorems. *Numerische Mathematik*, 2(1):137–141, 1960.

[9] R. Legenstein and W. Maass. Edge of chaos and prediction of computational performance for neural circuit models. *Neural networks*, 20(3):323–334, 2007.

[10] C. Gallicchio, A. Micheli, and L. Pedrelli. Fast spectral radius initialization for recurrent neural networks. In *Proceedings of INNSBDDL*, pages 380–390, 2019.