Analysis of Tensor-Based Image Segmentation using Echo State Networks

Charles Donkor[1], Emmanuel Sam[2], and Sebastian Basterrech[3]

[1] Department of Computer Science
University of Cape Coast,Cape Coast, Ghana
Cdonkor@knust.edu.gh
[2] School of Computing and Technology
Wisconsin International University College, Ghana
elsam@wiuc-ghana.edu.gh
[3] Department of Computer Science, Faculty of Electrical Engineering
Czech Technical University, Prague, Czech Republic.
Sebastian.Basterrech@fel.cvut.cz

Abstract. In pixel classification-based segmentation both the quality of the feature set and the pixel classification technique employed may influence the accuracy of the process. Motivated by the potential of Structural Tensors in the extraction of hidden information about texture between neighboring pixels and the computational capabilities of Echo State Network (ESN), this study proposes a tensor-based segmentation approach using the standard ESN. The pixel features of the image are initially extracted by incorporating Structural Tensors in order to enrich it with information about image texture. Then the resulting feature set is fed into ESN and the output is trained to classify unseen pixels from the testing set. The effect of the two main parameters that impact the accuracy of ESN: reservoir size and spectral radius, was also evaluated. The results are promising when compared to recent state-of-the-art segmentation approaches.

Keywords: Image segmentation, Image processing, Echo State Network, Support Vector Machine, Feature extraction

1 Introduction

An image segmentation technique consists of partitioning an image into different meaningful regions such that each region is homogeneous [1]. It forms an essential component of image processing, as it seeks to delineate the boundaries of various objects of interest in an image or parts of an image for further analysis. The delineation is usually based on image features such as colour, shape, texture or a mixture of these. But the process is accompanied by complexities such as the presence of noise, overlap between intensities of different objects,variation of contrast and weak edges [2]. Therefore, even though a number of segmentation techniques have been proposed, no single technique can be considered good for all images. The quality of image segmentation based on features created by structural Tensor has been analyzed recently in [3]. Structural Tensor (ST) is capable of extracting information on texture, which is within the neighborhood of a pixel. In [3], structural information extended with information about color, intensity, or a mixture of these features was found to be useful in achieving optimal color image segmentation outcome. Therefore this study chooses to adopt this method of feature extraction to test the quality of image segmentation using Echo State Network (ESN) [4].

ESN and a closely related approach known as Liquid State Machine (LSM) [5] were introduced in the early 2000s as alternative approach to gradient-descent based approaches for training Recurrent Neural Networks (RNNs). The computational approach introduced by these two independent but simultaneous models, which has recently become known as Reservoir Computing, demonstrates that RNN can still perform significantly well even when only a subset of the network weights are trained. ESN consists of a RNN of fixed random weights known as reservoir, and a supervised learning model called the readout. When driven by an input signal, the reservoir improves the linear separability of the input data using a high dimensional feature map, and preserves the nonlinear transformation of the input history in its internal states. A classification or prediction problem can then be solved by training only the weights of the readout structure using the collected reservoir activation states.Therefore the training procedure is fast, and avoids the problem of vanishing exploding gradient [6] introduced by gradient-descent methods. However, the reservoir is influenced by a number of global parameters which in turn impacts the accuracy of the model.The most relevant of these parameters are sparsity, spectral radius of the reservoir matrix, and reservoir size or the dimension of the reservoir matrix [7]. Guidelines on how these parameters can be tuned to guarantee that the state of the reservoir is suitable for good predictions can be found in [4, 8].

The internal reservoir activations provided by the dynamics of ESN reservior has been sufficient in solving many benchmark problems(e.g. [4, 9]) and practical problems such as time series predictions [8, 10]. In [2], the potential of the ESN reservoir to refine pixel features for colour image segmentation and the influence of the above mentioned parameters on the results have been investigated. In this work, the readout for the classification was realized with a Multi-Layer Perceptron (MLP) consisting of two hidden layers with 15 neurons. Besides, a tensor-based supervised classification method that uses Tucker decomposition to approximate the outputs from the ESN reservoir was proposed in [11], and numerical experiments carried out with spatiotemporal data outperforms the traditional method based on linear output weight, in terms of classification accuracy.

However, this study explores the potential of the standard ESN model with linear regression readout for achieving good image segmentation results when applied on Tensor-based feature set. To test the resilience of ESN to redundant and noised pixel attributes, no feature reduction techniques were applied. Finally, we evaluate the influence of spectral radius and reservoir size on the accuracy of image segmentation and compare the best result with the accuracy attained with

Support Vector Machine [12] (when applied on the same tensor-based feature set) as well as the accuracy of existing state of the art methods. The rest of this paper is organized as follows. Section 2.2 presents a description of Structural Tensor as well as ESN and its properties. 3 provides a description of the dataset and the experimental setup for this study. The results of the experiments and related discussions on how reservoir size and spectral radius impacts the quality of image segmentation are presented in Section 4. We conclude and present recommendation for future work in Section 5.

2 Background

This section consist of two parts. It presents a description of the concept of Structural Tensor, followed by a description of the ESN model and its properties.

2.1 Structural Tensor

Structural tensor allows for the extraction of valuable information on texture in addition to information on color and intensity [3]. For any given two dimensional image I, the structural tensor $\mathbf{ST}$ at a point $\mathbf{p_0}$ can be computed using the following formula:

$$\mathbf{ST}_{(\mathbf{p_0})} = G_{R_{(\mathbf{p_0})}}(\mathbf{DD}^T) \tag{1}$$

where $\mathbf{R}(\mathbf{p_0})$ is the compact nearest neighbourhood of $\mathbf{p_0}$, $G_{R_{(\mathbf{p_0})}}$ is an averaging operator in the region R, centered at a point $\mathbf{p_0}$, and D denotes an image gradient vector at each point $\mathbf{p}$ in R, computed as follows [3]:

$$\mathbf{D}(\mathbf{p}) = \begin{bmatrix} I_x(\mathbf{p}) \\ I_y(\mathbf{p}) \end{bmatrix} \tag{2}$$

where $I_x(\mathbf{p})$ and $I_y(\mathbf{p})$ are discrete spatial derivatives of I at point $\mathbf{p}$ in the x and y directions respectively. The averaging operator $G_{R_{(\mathbf{p_0})}}$, can simply be realized using a discrete binomial or Gaussian filter [13]. However, for more precise computations nonlinear anisotropic filter is used. Discussions on this can be found in [3]. After applying the filter G_R to average over a set of points in R, $\mathbf{ST}$ becomes a symmetric positive 2D matrix which elements describe average values of the gradient components in the neighborhood defined at the given point $\mathbf{p_0}$ [3]:

$$\mathbf{ST} = G_R\left(\begin{bmatrix} I_x \\ I_y \end{bmatrix} \begin{bmatrix} I_x I_y \end{bmatrix}\right) = \begin{bmatrix} I_x I_x & I_x I_y \\ I_y I_x & I_y I_y \end{bmatrix} = \begin{bmatrix} T_{xx} & T_{xy} \\ T_{xy} & T_{yy} \end{bmatrix} \tag{3}$$

The structural tensor $\mathbf{ST}$ does not only carry information on signal changes at a single point $\mathbf{p_0}$ but also at all points in the nearest neighborhood of $\mathbf{p_0}$. Therefore it conveys information on overlapping regions (i.e. image texture and local curvature) around $\mathbf{p}$. To include information on colour/intensity, the 2D vector $\mathbf{D}(\mathbf{p})$ can be extended to get the 3D vector $\mathbf{E}$:

$$\mathbf{E^T}(\mathbf{p}) = \begin{bmatrix} \mathbf{D^T} \ \mathbf{I}(\mathbf{p}) \end{bmatrix}^{\mathbf{T}} = \begin{bmatrix} I_x \ I_y \ I \end{bmatrix}^{\mathbf{T}} \tag{4}$$

By substituting 4 into 1, the Extended Structural Tensor(EST) is obtained as follows:

$$\mathbf{EST} = G_R(EE^T) = G_R\left(\begin{bmatrix} I_x \\ I_y I \end{bmatrix} \begin{bmatrix} I_x I_y I \end{bmatrix}\right) = \begin{bmatrix} I_x^2 & I_x I_y & I_x I \\ I_y I_x & I_y^2 & I_y I \\ I_x I & I_y I & I^2 \end{bmatrix} \tag{5}$$

Now **EST** contains average components of the gradient, besides average squared intensity signal, as well as mixed products of the gradient component and intensity [3]. To account for the components I_R, I_G, I_B of a colour image, **EST** can be obtained by replacing **D** in 1 with the following vector: the following vector **F**:

$$\mathbf{F^T}(\mathbf{p}) = \left[\mathbf{D^T}\ \mathbf{I}(\mathbf{p})\right]^{\mathbf{T}} = \left[I_x\ I_y\ I_R\ I_G\ I_B\right]^{\mathbf{T}} \tag{6}$$

The result is a positive definite symmetrical matrix, which contains fifteen components.

2.2 Description of Echo State Network

As shown in 1, the architecture of ESN consist of an input layer with p neurons connected to d hidden neurons, and an output layer with o neurons. The connections between these layers form two main structures: a reservoir structure and a readout structure. The reservoir (RNN) structure is defined by the tuple ($\mathbf{W}^{\text{in}}$,$\mathbf{W}^{\text{r}}$), where $\mathbf{W}^{\text{in}}$ and $\mathbf{W}^{\text{r}}$ are randomly generated input connection (input-to-hidden) and recurrent connection (hidden-to-hidden) weight matrices with dimensions $d \times (1 + p)$ and $d \times d$ respectively, and the readout consist of a hidden-to-output weight matrix $\mathbf{W}^{\text{out}}$ with dimensions $o \times (1+p+d)$, which are usually trained with linear regression model. The 1 accounts for the dimension of the first row of $\mathbf{W}^{\text{in}}$ and $\mathbf{W}^{\text{out}}$ which usually contains 1s corresponding to bias terms. When p is driven by an input signal $\mathbf{s}(t)$ from an input space $\mathbb{R}^p$ at any

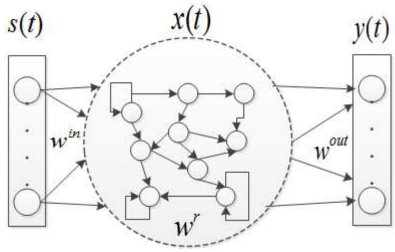


Fig. 1: Architecture of a standard ESN model.

time t, the reservoir uses $\mathbf{W}^{\text{in}}$ as a high dimensional feature map to transform the signal into a larger space $\mathbb{R}^d$ with $p \ll d$, and then memorizes the nonlinear transformations of the input history in its internal states.

Given a training set composed of input signal $\mathbf{s}(t) \in \mathbb{R}^p$, the reservoir updates its internal states $\mathbf{x}(t) = (x_1(t), ..., x_N(t))$ according to the following formula:

$$\mathbf{x}(t) = g_h(\mathbf{s}(t), \mathbf{x}(t-1), \mathbf{W}^{\text{in}}, \mathbf{W}^{\text{r}}). \tag{7}$$

Next, the parametric function shown below uses the actual reservoir states to execute the model output:

$$\hat{\mathbf{y}}(t) = g_o(\mathbf{x}(t), \mathbf{W}^{\text{out}}), \tag{8}$$

where $g_o(\cdot)$ is an activation function with parameters in $\mathbf{W}^{\text{out}}$.

Although in the standard ESN model there are no connections between the input and readouts neurons [4], another readout form is the following [8]:

$$\hat{\mathbf{y}}(t) = g_o(\mathbf{s}(t), \mathbf{x}(t), \mathbf{W}^{\text{out}}). \tag{9}$$

In this study, we used hyperbolic tangent $\tanh(\cdot)$ as the activation function $g_h(\cdot)$, and so the dynamics was computed as:

$$\mathbf{x}(t) = \tanh(\mathbf{W}^{\text{in}}\mathbf{s}(t) + \mathbf{W}^{\text{r}}\mathbf{x}(t-1)). \tag{10}$$

The output of the model, $\hat{\mathbf{y}}(t)$ at a given time t is computed as follows:

$$\hat{\mathbf{y}}(t) = \mathbf{W}^{\text{out}}[\mathbf{s}(t); \mathbf{x}(t)], \tag{11}$$

where $[\cdot;\cdot]$ denotes a vector concatenation operation.

3 Methodology

In this section, we explain the dataset and the experimental setup used to evaluate the usefulness and effectiveness of ESN for tensor-based colour image segmentation.

3.1 Data Description

The datasets employed in this study to analyze the quality of tensor-based image segmentation using ESN were the original datasets extracted by a proposed Tensor-Based Image Segmentation Algorithm (TBISA) in [3] for a similar study using other classifiers. The datasets were based on Berkley Segmentation Benchmark images [14], which consisted of seven images identified as '35058', '41033', '66053', '69040', '134052', '161062',and '326038'. These images consisted of masks for two and three class detection problems. However, this study concentrated on the masks for two class detection problems. The masks were defined by [3] based on the original segmentation contour published on source page.

The EST feature set is made up of fifteen features which include tensor information on a pixel and its neighborhood, as well as mixed products of these. Besides, it has a corresponding class label that indicates whether the pixel belongs to an object or a background. Detailed description of the extraction process, and the Tensor-Based Image Segmentation Algorithm (TBISA) employed can be found in [3].

3.2 Experimental Setup

First of all, the EST datasets for all the seven images were combined into one dataset. The values of attributes which form each feature vector were then normalized to lie between 0 and 1, and the pixels were randomized. The resulting data was divided into two subsets: 80% was used to train the ESN and the remaining 20% was used to test it. We experimented with two different classifiers: a standard ESN where $\mathbf{W}^{\text{in}}$ and $\mathbf{W}^{\text{r}}$ are randomly initialized with uniformly distributed weights in the range $[-0.5, 0.5]$, and SVM classifier which uses a Gaussian radial basis (RBF) kernel function. In order to test the effect of spectral radius and reservoir size on the effectiveness of ESN as a classification algorithm, the reservoir was configured with different combinations of the following reservoir sizes d and spectral radius ρ: $d = \{150, 200, 300, 400, 500, 600\}$ and $\rho = \{0.1, 0.3, 0.5, 0.99\}$. The selection of these range of values were informed by previous researches (e.g. [15, 2]) which considered the influence these parameters on the performance of ESN. The reservoir states were updated with a leaking rate of 0.3 and the output weights were estimated by setting the regularization factor in the linear regression model to 1×10^{-2}. The classification accuracy resulting from the selected pair of parameters (d, ρ), was estimated using Accuracy. The experiment for each combination of parameters was repeated five times and the average accuracy was recorded.

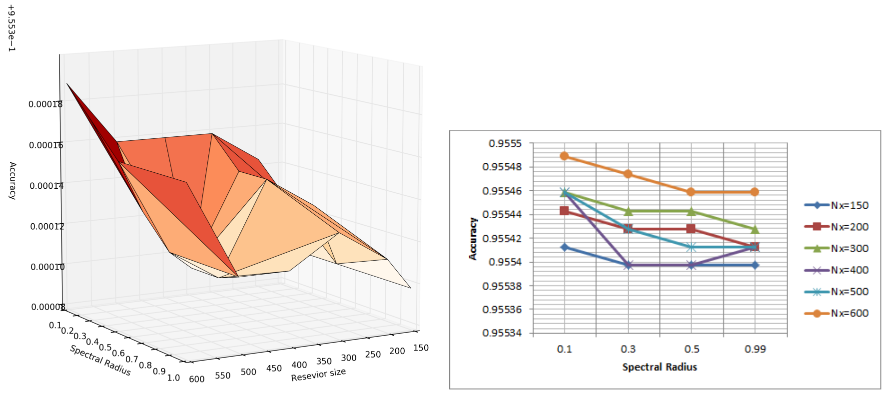
4 Results and Discussion

The mean and standard deviation of all the accuracies that resulted from the various combinations of d and ρ were 0.955431026 and 0.000026323 respectively. Table 1 below shows the maximum and minimum accuracies as well as the values of d and ρ that led to each accuracy. These results imply that, all things being

Table 1: Classification Performance of ESN when applied on Extended Structural Tensor Feature Set.

	Accuracy	**Reservoir Size(d)**	**Spectral Radius(ρ)**
Maximum	0.955488667	600	0.1
Minimum	0.955397455	400	0.3

equal, changes in spectral radius and reservoir size leads to a fairly significant change in accuracy. The trend for the changes realized is depicted (in both 3D and 2D format) in figure 2 below.It can be observed that, in most cases, an increase in the value of ρ caused a slight corresponding improvement in classification accuracy. Besides, for each spectral radius, an increase in d leads to a positive change in accuracy in virtually all cases. Hence the best accuracy, as shown in table1, was attained with a spectral radius of 0.1 and reservoir size of 600. This confirms the observation of [2] about spectral radius in a similar study where ESN was used for feature selection and MLP was used as readout for classification, and a spectral radius of 2.0 or less was proposed as the best choice for good quality colour image segmentation. Our finding for the best reservoir size is, however, contrary to what was suggested in [2], as $d > 400$ rather appear to be a good choice of reservoir dimension for a good image segmentation quality. These findings served as motivation for further experiments to ascertain if classification accuracy can get any better when d is increased beyond 600 neurons and ρ is decreased below 0.1. Therefore combinations of the following values of d and ρ were tested: $d = 700, 800, 900, 1000$ and $\rho = 0.01, 0.03, 0.05, 0.09$. As depicted in figure 3 below, a steady increase in accuracy was realized as d increases. The effect of ρ, however, became unpredictable and did not have as much effect as d. Based on this, we selected 0.9565224 attained with $d = 1000$ and $\rho = 0.01$ as



(a) (b)

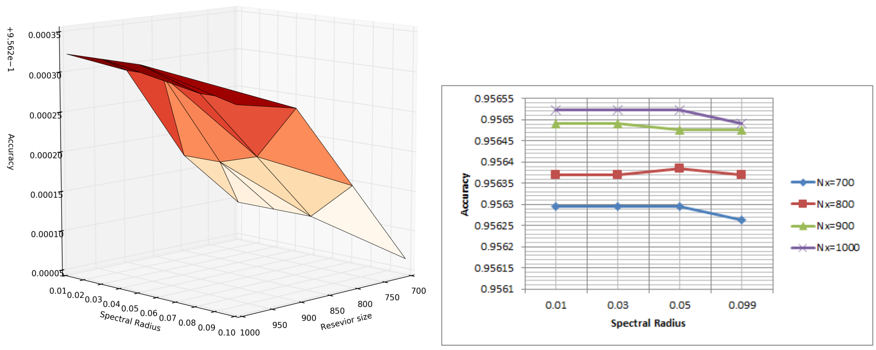
Fig. 2: Effect of a range of d and ρ on the Accuracy of Tensor-based Image Classification using ESN. d ranges from 150 - 600 and ρ ranges from 0.1 - 0.99.

the best performance of ESN for color image segmentation as far as this study is concerned.

4.1 Comparing ESN and other Techniques

Table 2 below shows the best performance of ESN obtained through the experiments explained above; the accuracy of SVM when applied on the same EST feature set; and the results obtained in related studies using other approaches.

It can be seen that the segmentation accuracy of ESN is relatively better than the average performance of SVM as far as this study is concerned. Although our tensor-based approach did not involve any form of feature selection, as was done in some related studies (e.g. [2, 11]), the performance of ESN was promising. This demonstrates the resilience of ESN to redundant and noised data. It is also worth mentioning that in [3], the accuracy was estimated over each image dataset; the pixels used for training were also used for testing. But this study took a different approach in order to test the ability of ESN to classify unseen pixels and also to make the model insusceptible to over-fitting: pixels used for testing were different from those used for training. When compared to the classification result attained



(a) (b)

Fig. 3: Effect of a range of d and ρ on the Accuracy of Tensor-based Image Classification using ESN. d ranges from 700 - 1000 and ρ ranges from 0.01 - 0.099.

by the tensor-based method proposed in [11], it can be concluded that the result of approximating the reservoir outputs using a Tucker decomposition may be similar to that of using raw reservoir outputs based on EST feature set.

Table 2: Comparison of ESN to SVM and some existing State of the Art Classifiers.

Algorithm	Accuracy	Dataset
ESN	0.9565224	All images combined*
SVM	0.86106	All images combined*
Random Tree(RT) [3]	0.9322	'326038'*
MLP [3]	0.9938	'35058'*
MLP [2]	0.926	SSDS dataset [2]
*'35058','41033','66053','69040','134052','161062','326038'		

5 Conclusions and Future Work

In this paper, we have investigated the usefulness of ESN for color image segmentation when applied on Extended Structural Tensor features. We also explored the effect of the two main parameters of Echo State Network (ESN): spectral radius and reservoir size on image segmentation accuracy, and we compared our result to related state of the art methods. From our experiments, it can be inferred that both the reservoir size and spectral radius have fairly significant effects on pixel classification accuracy. After an initial set of experiments with reservoir size ranging from 150 to 600 and spectral radius ranging from 0.1 to 0.99, the best choice of spectral radius and reservoir size were 0.1 and 600 neurons respectively. Further tests with $\rho < 0.1$ and $d > 600$ led to further improvement in accuracy. Based on this, 0.9565224 attained with $d = 1000$ and $\rho = 0.01$

was chosen as the best performance of ESN for color image segmentation as far as this study is concerned Tensor-based segmentation accuracy using ESN was comparatively better than SVM, and showed promising results when compared with recent state-of-the-art segmentation approaches. However, in future work, we plan to analyze the accuracy of other types of ESN models, and we expect to apply the NN and Structural Tensor feature set refined through various feature selection processes.

Acknowledgment

This work has been supported by the Czech Science Foundation (GAČR) under research project No. 18-18858S, and the authors acknowledge the support of the OP VVV MEYS funded project CZ.02.1.01/0.0/0.0/16_019/0000765 "Research Center for Informatics".

References

1. N. R. Pal and S. K. Pal, "A review on image segmentation techniques," *Pattern Recognition*, vol. 26, pp. 1274–1294, 1993.
2. A. Souahlia, A. Belatreche, A. Benyettou, and K. Curran, "An experimental evaluation of echo state network for colour image segmentation," *2016 International Joint Conference on Neural Networks (IJCNN)*, pp. 1143–1150, 2016. [Online]. Available: http://ieeexplore.ieee.org/document/7727326/
3. K. Jackowski and B. Cyganek, "A learning-based colour image segmentation with extended and compact structural tensor feature representation," *Pattern Analysis and Applications*, vol. 20, pp. 401–414, 2017.
4. H. Jaeger, "The "echo state" approach to analysing and training recurrent neural networks," German National Research Center for Information Technology, Tech. Rep. 148, 2001.
5. W. Maass, T. Natschläger, and H. Markram, "Real-time computing without stable states: a new framework for a neural computation based on perturbations," *Neural Computation*, vol. 14, pp. 2531–2560, november 2002.
6. Y. Bengio, P. Simard, and P. Frasconi, "Learning long-term dependencies with gradient descent is difficult," *Neural Networks, IEEE Transactions on*, vol. 5, no. 2, pp. 157–166, 1994.
7. M. Lukoševičius, "A Practical Guide to Applying Echo State Networks," in *Neural Networks: Tricks of the Trade*, ser. Lecture Notes in Computer Science, G. Montavon, G. Orr, and K.-R. Müller, Eds. Springer Berlin Heidelberg, 2012, vol. 7700, pp. 659–686. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-35289-8_36
8. M. Lukoševičius and H. Jaeger, "Reservoir Computing Approaches to Recurrent Neural Network Training," *Computer Science Review*, vol. 3, pp. 127–149, 2009.
9. H. Jaeger and H. Haas, "Harnessing Nonlinearity: Predicting Chaotic Systems and Saving Energy in Wireless Communication," *Science*, vol. 304, no. 5667, pp. 78–80, 2004.
10. S. Basterrech and G. Rubino, "Echo State Queueing Networks: A Combination of Reservoir Computing and Random Neural Networks," *Probability in the Engineering and Informational Sciences*, vol. 31, pp. 457–476, October 2017. [Online]. Available: https://doi.org/10.1017/S0269964817000110

11. A. Prater, "Classification via Tensor Decompositions of Echo State Networks," in *2017 IEEE Symposium Series on Computational Intelligence (SSCI)*, 2017, pp. 1–8.
12. H. Drucker, C. Burges, L. Kaufman, A. Smola, and V. Vapnik, "Support Vector Regression Machines," *Neural Information Processing Systems*, vol. 1, pp. 155–161, 1996.
13. L. D. Lathauwer, *Signal Processing Based on Multilinear Algebra*. Katholieke Universiteit Leuven Faculteit der Toegepaste Wetenschappen Department Elektrotechniek, 1997.
14. D. Martin, C. Fowlkes, D. Tal, and J. Malik, "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics," in *Proc. 8th Int'l Conf. Computer Vision*, vol. 2, July 2001, pp. 416–423.
15. S. Basterrech, "Empirical analysis of the necessary and sufficient conditions of the echo state property," in *2017 International Joint Conference on Neural Networks, IJCNN 2017, Anchorage, AK, USA, May 14-19, 2017*, 2017, pp. 888–896.