NOLTA, IEICE

Paper

Reservoir-based convolution

Yuichiro Tanaka [1 a)] *and Hakaru Tamukoh* [1,2]

[1] *Research Center for Neuromorphic AI Hardware, Kyushu Institute of Technology 2-4 Hibikino, Wakamatsu, Kitakyushu 808-0196, Japan*

[2] *Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology 2-4 Hibikino, Wakamatsu, Kitakyushu 808-0196, Japan*

[a)] *tanaka-yuichiro@brain.kyutech.ac.jp*

Received October 18, 2021; Revised December 18, 2021; Published April 1, 2022

Abstract: Reservoir computing (RC) has attracted attention and has been used in many applications because of its low training cost. Multiple studies using RC for image recognition have been proposed, and some have achieved accuracy rates of greater than 99% on the MNIST dataset. For the Fashion-MNIST and CIFAR-10 datasets, however, they have not yet achieved high accuracy. This study proposes a novel convolutional neural network based on RC that can be optimized by ridge regression rather than back-propagation. The reservoir-based network has multiple reservoirs with various leak rates to extract features with various spatial frequencies from the inputs. The experimental results show that the performance of the proposed model achieves higher accuracy rates in the mentioned datasets compared with those of other reservoir-based image recognition approaches.

Key Words: convolutional neural network, echo state network, image recognition, neural network, reservoir computing.

1. Introduction

Reservoir computing (RC) [1] is a type of recurrent neural network in which weight connections in a hidden layer do not have plasticity, whereas weights between the hidden and output layers do. This is done so that the training costs will be lower than those of deep neural networks when facing a large number of plastic weight connections trained by back-propagation and stochastic gradient descent (SGD) methods.Owing to the low training cost, RC has attracted attention and has been used in several applications in recent years.

Image recognition is one of several applications of RC [2–6]. Shaetti *et al.* [2] investigated the abilities of RC (i.e., echo-state networks (ESNs)) [1, 7] on image recognition tasks and showed that an ESN with 4,000 nodes in the reservoir achieved an accuracy rate of 99.07% on the MNIST dataset [8] by applying appropriate preprocessing to the dataset. Tong and Tanaka [3] proposed a model combining a convolutional neural network (CNN) [8] and RC, where the convolution and pooling layers worked as dimension reducers of input data, and the RC received and processed the compressed data to classify the input. The model achieved an accuracy rate of 99.25% on the MNIST dataset. Yonemura and Katori [4] investigated the relationship between the number of training parameters and the accuracy

Nonlinear Theory and Its Applications, IEICE, vol. 2, no. 3, pp. 1101–1106 DOI: 10.1587/nolta.2.1101

of Tong's model. Their study showed that a model with 40K training parameters achieved 98.71% on the MNIST dataset, whereas the original work [3] required 90K training parameters. An *et al.* [5] proposed a model combining a deep neural network and a delay feedback reservoir (DFR) [9] called deep-DFR, which achieved an accuracy rate of 99.03% on the MNIST dataset. Although the deep-DFR model is a reservoir-based approach, the model was optimized by the SGD, which required a high training cost. Velichko [6] proposed a neural network that used filters based on logistic mapping called LogNNet, which achieved an accuracy rate of 96.3% on the MNIST dataset.

As mentioned, some studies have achieved accuracy rates of greater than 99% on the MNIST dataset. However, such high accuracy rates have not been achieved on the Fashion-MNIST [10] and CIFAR-10 [11] datasets. For example, Tong's model [4] achieved an accuracy rate of 86.27% on the Fashion-MNIST dataset, and the deep-DFR model [5] achieved an accuracy rate of 60.57% on the CIFAR-10 dataset. This study aims to realize a reservoir-based neural network that achieves high accuracy rates on the Fashion-MNIST and CIFAR-10 datasets with low training costs. Hence, we propose a novel reservoir-based convolutional operation.

This paper is organized as follows. Section 2 describes the proposed reservoir-based convolutional operation. Section 3 describes the experimental settings and the results. Section 4 provides a discussion, and Section 5 concludes the paper.

2. Proposed method

Figure 1 shows the proposed reservoir-based convolution layer that receives a region of interest (ROI) clipped from input feature maps. The ROI is shifted using a stride size from the top-left to the bottom-right of the input in the same manner as the CNNs. The ROIs, whose channel, height, and width are C, K_h, and K_w, respectively, are fed into reservoirs in the layer. The layer has two streams. One is a horizontal stream that receives a $C \times K_h$ dimensional vector, $\boldsymbol{u}^{\mathrm{hor}}(t)$, where t is a time step $(0 \leq t \leq K_w)$, and the other is a vertical stream that receives a $C \times K_w$ dimensional vector, $\boldsymbol{u}^{\mathrm{ver}}(t)$ $(0 \leq t \leq K_h)$. The reservoir state, $\boldsymbol{x}(t) \in \mathbb{R}^R$ ($\boldsymbol{x}^{\mathrm{hor}}(t)$ in the horizontal stream and $\boldsymbol{x}^{\mathrm{ver}}(t)$ in the vertical stream), is updated using Eq. 1. Note that $\boldsymbol{x}(0) = \mathbf{0}$.

$$\boldsymbol{x}(t+1) = f\{(1-\delta)\boldsymbol{x}(t) + \delta(W_{\mathrm{in}}\boldsymbol{u}(t) + W_{\mathrm{rec}}\boldsymbol{x}(t))\}. \tag{1}$$

W_{in} and W_{rec} in the equation are a weight matrix between the input and the reservoir and a recurrent connection matrix in the reservoir, respectively. These matrices are randomly initialized and are not

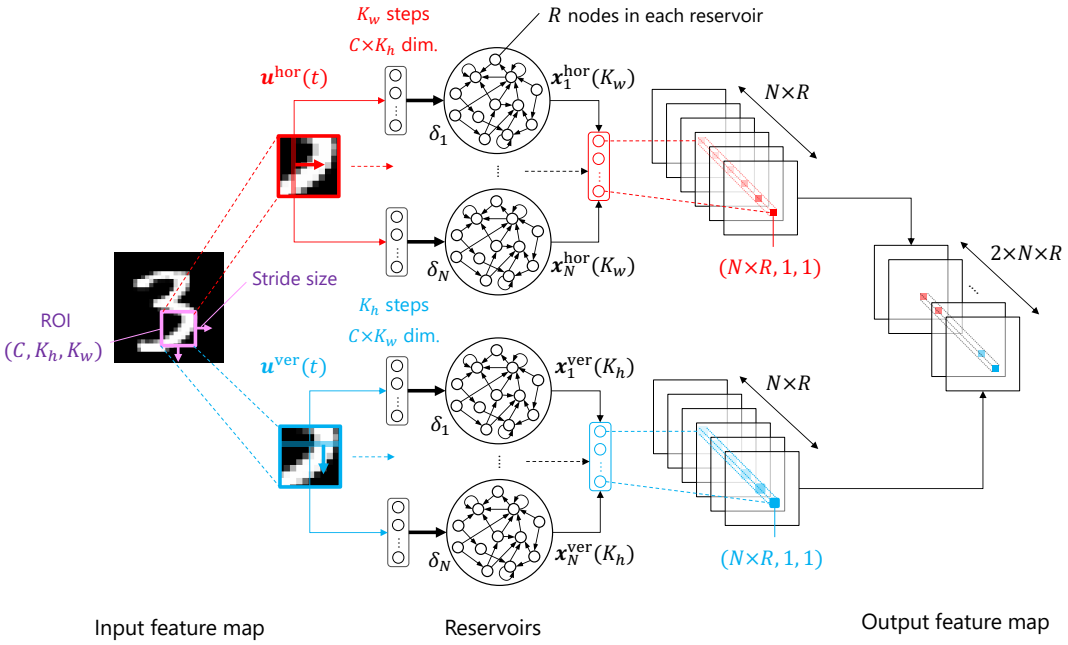


Fig. 1. Reservoir-based convolution layer.

updated during training. Note that the connection density of the matrix, W_{rec}, is set as p $(0 \leq p \leq 1)$, and the matrix is scaled to satisfy the echo state property [12]. f indicates a nonlinear activation function; tanh is used in this study. δ $(0 < \delta < 1)$ indicates the leak rate that controls the updating speed of the reservoir state.

The reservoir-based convolution layer has multiple reservoirs (N reservoirs in each stream) having various leak rates. Hence, to extract features with various spatial frequencies from the input feature maps, a reservoir having a low leak rate slowly updates its reservoir state and extracts features with low spatial frequency and a reservoir with high leak rate extracts features having high spatial frequency. After feeding the ROI to the reservoirs, the reservoir states in each stream are concatenated to form an $N \times R$ channel pixel, and the $N \times R$ channel pixels of both streams are concatenated to form a $2 \times N \times R$ channel pixel for the output feature maps.

3. Experiments

3.1 Image classification

We constructed a neural network that includes the proposed reservoir-based convolution layers, as shown in Fig. 2 and Table I. We evaluated network performance using the MNIST, Fashion-MNIST, and CIFAR-10 datasets, where both MNIST and Fashion-MNIST datasets included 60,000 training images and 10,000 test images, and the CIFAR-10 dataset included 50,000 training images and 10,000 test images. The parameters of the reservoirs were set as $N = 5$, $R = 12$, and $p = 0.5$ for the first reservoir-based convolution layer, and $N = 5$, $R = 30$, and $p = 1.0$ for the second. The leak rate of the i-th reservoir in a layer is $\delta_i = 0.8 \times (i-1)/(N-1) + 0.1$, such that $\delta_i = 0.1, 0.3, 0.5, 0.7, 0.9$ in the case of $N = 5$. During training, we fed the training data into the network and updated the weight connections only in the fully connected layer using ridge regression, $W_{\mathrm{fc}} = (X^\top X + \lambda I)^{-1} X^\top Y$, where W_{fc}, X, and Y are the weight matrix in the fully connected layer, input vectors of the layer, and target vectors of the layer, respectively. The target vector of each training data was given as a one-hot vector corresponding to the data label. $\lambda(> 0)$ is the regularization strength, and I is the identity matrix.

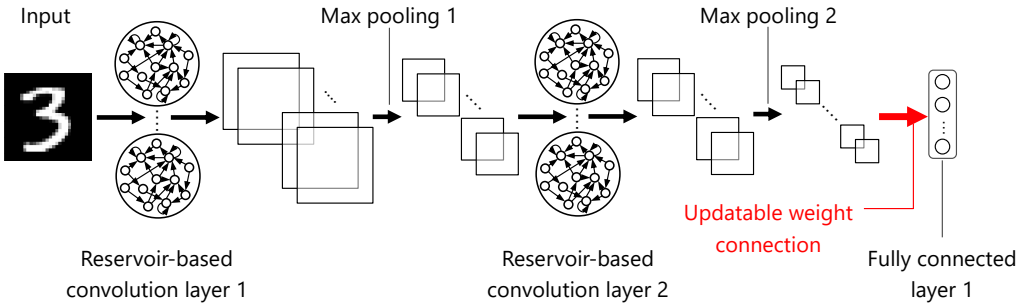


Fig. 2. Reservoir-based convolutional neural network.

Table I. Structure of the network.

Layer name	Parameters
Reservoir-based convolution 1	Kernel size: 3, Stride size: 1
Max pooling 1	Kernel size: 2, Stride size: 2
Reservoir-based convolution 2	Kernel size: 3, Stride size: 1
Max pooling 2	Kernel size: 2, Stride size: 2
Fully connected 1	Output size: 10

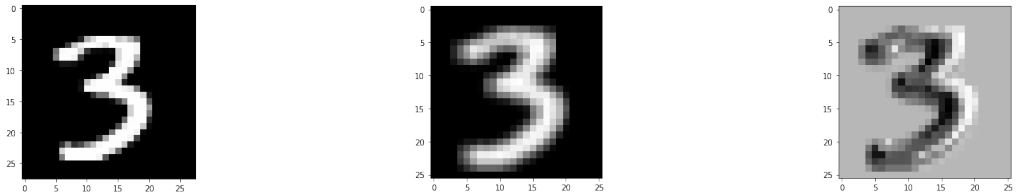
After training, we fed the test data into the trained network to calculate the accuracy rates. Table II shows a comparison of accuracy rates for the MNIST, Fashion-MNIST, and CIFAR-10 datasets comparing this study to other reservoir-based image recognition approaches. The proposed network did not achieve an accuracy rate of 99% for the MNIST dataset, whereas it outperformed other approaches for the Fashion-MNIST and CIFAR-10 datasets.

Table II. Accuracy rates in image recognition tasks.

	MNIST	Fashion-MNIST	CIFAR-10
[2]	99.07%	N/A	N/A
[3]	99.25%	N/A	N/A
[4]	98.71%	86.27%	N/A
[5]	99.03%	N/A	60.57%
[6]	96.30%	N/A	N/A
This study	98.38%	91.04%	64.49%

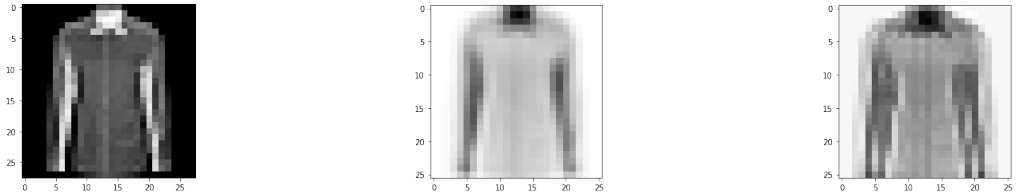
3.2 Feature extraction

To verify that the proposed network extracts features with various spatial frequencies as inputs using multiple reservoirs, we fed a handwritten digit image from the MNIST dataset, a clothes image from the Fashion-MNIST dataset, and a horse image from the CIFAR-10 dataset into the proposed network. We obtained output feature maps from reservoirs having leak rates of $\delta = 0.1$ and 0.9 in the first reservoir-based convolution layer. Figures 3, 4, and 5 show the input handwritten digit, clothes, and horse images, and the output feature maps, respectively. In the case of the MNIST dataset, the output from the reservoir with $\delta = 0.1$, which slowly updated its state, was blurred while the reservoir with $\delta = 0.9$ that quickly updated its state extracted the edge of the handwritten line. In the cases of the Fashion-MNIST and CIFAR-10 datasets, the reservoir with $\delta = 0.1$ extracted outlines of clothes and horses, whereas the reservoir with $\delta = 0.9$ extracted the textures.



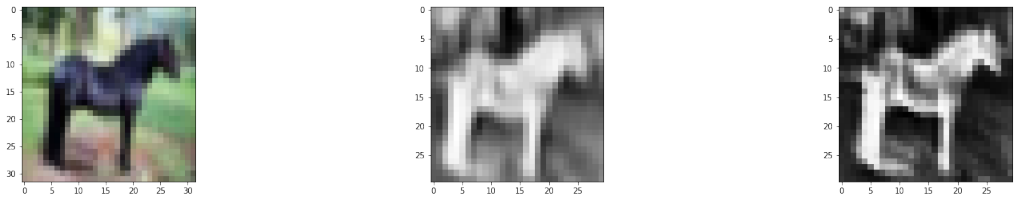
(a) Input (b) Output ($\delta = 0.1$) (c) Output ($\delta = 0.9$)

Fig. 3. Input image from MNIST dataset and output feature maps from the reservoir-based convolution layer.



(a) Input (b) Output ($\delta = 0.1$) (c) Output ($\delta = 0.9$)

Fig. 4. Input image from Fashion-MNIST dataset and output feature maps from the reservoir-based convolution layer.



(a) Input (b) Output ($\delta = 0.1$) (c) Output ($\delta = 0.9$)

Fig. 5. Input image from CIFAR-10 dataset and output feature maps from the reservoir-based convolution layer.

4. Discussion

From the viewpoint of biology, a reservoir-based CNN has several similarities with the human brain. A CNN structure is based on cells in the visual cortex: the partial connections using filters (images receptive fields) in the convolution layer correspond to the simple cells, and the pooling corresponds to the complex cells. This is also true for the proposed network because the reservoir-based convolution layer also has image receptive fields, and pooling is used in the network. However, CNNs are trained using the back-propagation method, which is not biologically plausible [14]. By contrast, the proposed network is trained without using back-propagation, and therefore it would be more biologically plausible than CNNs. Moreover, several studies have proposed cortex models using reservoirs, e.g., Katori proposed a prefrontal cortex model [15] and Yonemura and Katori proposed a cortex model integrating visual and auditory stimuli [16]. Therefore, the reservoir can be viewed as a cortex model, and the proposed network can be seen as one form of this type of model.

The reason why the proposed network outperformed the other reservoir-based approaches for the Fashion-MNIST and CIFAR-10 datasets, as shown in Table II, relied on multiple reservoirs with several leak rates in the reservoir-based convolution layers extracting various features of images that were important for recognition, as shown in Figs. 4 and 5. However, the proposed network did not achieve an accuracy rate of 99% for the MNIST dataset because images in the MNIST dataset did not have various spatial frequencies compared with the Fashion-MNIST and CIFAR-10 datasets, indicating that multiple reservoirs were not effective in the dataset.

The proposed network has a high accuracy rate for the Fashion-MNIST and CIFAR-10 datasets, but its training cost is also low. The ESN that achieves an accuracy rate of 99.07% for MNIST dataset [2] and Yonemura's network, which achieved accuracy rates of 98.71% for the MNIST dataset and 86.27% for the Fashion-MNIST dataset [4], required 40K training parameters. Moreover, Tong's model for the MNIST dataset required 90K training parameters [3]. The proposed network required 75,010 training parameters for the MNIST and Fashion-MNIST datasets (108,010 parameters for the CIFAR-10 dataset). Although the number of training parameters for the MNIST and Fashion-MNIST datasets of the proposed network was larger than those of [2] and [16], it was smaller than that of [3].

Compared with other reservoir-based approaches for image recognition, the proposed reservoir-based convolution layer requires a smaller memory capacity [13] because the reservoirs in the layer receive ROIs from the input feature maps, and three time steps are required to feed an ROI into the reservoirs in this study. Therefore, using physical RCs [17, 18] for the proposed network is possible, even if the physical RCs have smaller memory capacities compared with ESNs. Therefore, a low power implementation of the proposed network using the physical RCs is expected.

5. Conclusions

We proposed a reservoir-based convolutional operation and conducted image recognition tasks using the MNIST, Fashion-MNIST, and CIFAR-10 datasets with a neural network that included the reservoir-based convolution layers. During training, only the weights of the fully connected layer were updated by ridge regression, whereas other weights were not updated, such that the network required lower training cost than deep neural networks optimized by the back-propagation method. As shown in Table II, the proposed network outperformed other reservoir-based approaches for the Fashion-MNIST and CIFAR-10 datasets because multiple reservoirs with several leak rates extracted various features of images that were important for recognition, as shown in Figs. 4 and 5.

Although the proposed network outperformed other reservoir-based approaches, it still does not achieve state-of-the-art results. Byerly *et al.* achieved an accuracy rate of 99.87% for the MNIST dataset [19], Tanveer *et al.* achieved an accuracy rate of 96.91% for the Fashion-MNIST dataset [20], and Dosovitskiy *et al.* achieved an accuracy rate of 99.5% for the CIFAR-10 dataset [21]. We plan to make the structure of the proposed network deeper to improve its accuracy in future work.

Acknowledgments

This paper is based on results obtained from a project, JPNP16007, commissioned by the New Energy and Industrial Technology Development Organization (NEDO) and supported by JSPS KAKENHI

Grant Number 21K21318.

References

[1] H. Jaeger, "Tutorial on training recurrent neural networks, covering BPPT, RTRL, EKF and the "echo state network" approach," *GMD Report*, vol. 159, October 2002.

[2] N. Schaetti, M. Salomon, and R. Couturier, "Echo state networks-based reservoir computing for MNIST handwritten digits recognition," *International Conference on Computational Science and Engineering*, August 2016.

[3] Z. Tong and G. Tanaka, "Reservoir computing with untrained convolutional neural networks for image recognition," *2018 24th International Conference on Pattern Recognition*, pp. 1289-1294, August 2018.

[4] Y. Yonemura and Y. Katori, "Image recognition model based on convolutional reservoir computing," *The 34th Annual Conference of the Japanese Society for Artificial Intelligence*, June 2020.

[5] Q. An, K. Bai, L. Liu, F. Shen, and Y. Yi, "A unified information perceptron using deep reservoir computing," *Computers and Electrical Engineering*, vol. 85, July 2020.

[6] A. Velichko, "Neural network for low-memory IoT devices and MNIST image recognition using kernels based on logistic map," *Electronics*, vol. 9, no. 9, September 2020.

[7] H. Jaeger, "The "echo state" approach to analysing and training recurrent neural networks-with an erratum note," *GMD Report*, vol. 148, January 2001.

[8] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278-2324, November 1998.

[9] L. Appeltant, M.C. Soriano, G. Van der Sande, J. Danckaert, S. Massar, J. Dambre, B. Schrauwen, C.R. Mirasso, and I. Fischer, "Information processing using a single dynamical node as complex system," *Nature Communications*. vol. 2, no. 468, 2011.

[10] H. Xiao, K. Rasul, and R. Vollgraf, "Fashion-MNIST: a novel image dataset for benchmarking machine learning algorithms," arXiv:1708.07747, August 2017.

[11] A. Krizhevsky, "Learning multiple layers of features from tiny images," April, 2009.

[12] I.B. Yildiz, H. Jaeger, and S.J. Kiebel, "Re-visiting the echo state property," *Neural Networks*, vol. 35, pp. 1-9, 2012.

[13] H. Jaeger, "Short term memory in echo state networks," *GMD Report*, vol. 152, March 2002.

[14] T. Shinozaki, "Biologically motivated learning method for deep neural networks using hierarchical competitive learning," *Neural Networks*, vol. 144, pp. 271-278, 2021.

[15] Y. Katori, "Network model for dynamics of perception with reservoir computing and predictive coding," *Advances in Cognitive Neurodynamics (VI)*, pp. 89-95, 2018.

[16] Y. Yonemura and Y. Katori, "Multi-modal processing of visual and auditory signals on network model based on predictive coding and reservoir computing," 2020 International Symposium on Nonlinear Theory and Its Applications, pp. 209-212, 2020.

[17] K. Nakajima, "Physical reservoir computing—an introductory perspective," *Japanese Journal of Applied Physics*, vol. 59, no. 6, May 2020.

[18] Y. Usami, B. van de Ven, D. G. Mathew, T. Chen, T. Kotooka, Y. Kawashima, Y. Tanaka, Y. Otsuka, H. Ohoyama, H. Tamukoh, H. Tanaka, W. G. van der Wiel, and T. Matsumoto, "In-Materio Reservoir Computing in a Sulfonated Polyaniline Network," *Advanced Material*, September 2021,

[19] A. Byerly, T. Kalganova, I. Dear, "No routing needed between capsules," arXiv:2001.09136, January 2020.

[20] M.S. Tanveer, M.U.K. Khan, C.M. Kyung, "Fine-tuning DARTS for image classification," arXiv:2006.09042, June 2020.

[21] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An image is worth 16x16 words: transformers for image recognition at scale," *International Conference on Learning Representations*, 2021.