ESANN 2012 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 25-27 April 2012, i6doc.com publ., ISBN 978-2-87419-049-0. Available from http://www.i6doc.com/en/livre/?GCOI=28001100967420.

Complex Valued Artificial Recurrent Neural Network as a Novel Approach to Model the Perceptual Binding Problem

Alexey Minin[1,3], Alois Knoll[1] and Hans-Georg Zimmermann[2]

[1] Technische Universitaet Muenchen – Robotics and Embedded Systems, Boltzmannstraße 3, 85748 Garching bei Muenchen, Germany
[2] Siemens AG – Corporate Technology, 81730 Muenchen, Otto-Hahn Ring 6, Germany
[3] Siemens LLC – Corporate Technology, 191186 St. Petersburg, Volynskiy lane 3 of. 905, Russia.

Abstract. In this paper we suggest a new model for solving the binding problem by introducing complex-valued recurrent networks. These networks can represent sinusoidal oscillations and their phase, i.e., they can model the binding problem of neuronal assemblies by adjusting the relative phase of the oscillations of different feature detectors. As feature examples, we use color and shape – but the network would also function with any combination of other features. The suggested network architecture performs image generalization but can also be used as an image memory. The information about object color is represented in the phase of the network weights, while the spatial distribution of the neurons codes represent the object's shape. We will show that the architecture can generalize object shapes and recognize object color with very low computational overhead.

1 The Binding Problem

Brains are permanently confronted with the problem of classifying objects into meaningful groups, e.g., the object class of "cars", even though the objects can be of different size, color, shape. The binding hypothesis is as follows: there are groups of neurons that are sensitive to a certain colors and other groups that react to certain shapes (e.g., typical shapes that are associated with cars). When a human sees a green car, both the neurons responsible for the color green, and those neurons reacting to the car-specific shapes start oscillating. After some time both align their oscillation phase, which means that they react to the same object. Thus, in the perception of the human, the car is associated with the color green. More details are presented in works [1-4].

2 The Complex Valued Recurrent Neural Network

The basis of our work is the "recurrent open network" – a system that is driven not only by its internal processes but also by external inputs. The Complex-Valued Recurrent Neural Network (CVRNN) is an extension of the Real Valued Recurrent Neural Network (RVRNN) [5]. In the sequel we use CVRNNs that comply with the following general state space model:

ESANN 2012 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence
and Machine Learning. Bruges (Belgium), 25-27 April 2012, i6doc.com publ., ISBN 978-2-87419-049-0.
Available from http://www.i6doc.com/en/livre/?GCOI=28001100967420.

$$\begin{cases} s_{t+1} = \tanh\left(As_t + Bu_t\right), t = 1..ns \\ y_t = \tanh\left(Cs_t\right), t = 1..ns \end{cases}$$
$$E = \frac{1}{T}\sum_{t=1}^{ns}\left(y_t - y_t^d\right)\left(y_t - y_t^d\right)^* \quad (1)$$

Here E is an error function (the asterisk above the values denotes the complex conjugate) with E being the real-valued error. Fig. 1 illustrates the network architecture, i.e., the connections between inputs, outputs and nodes.

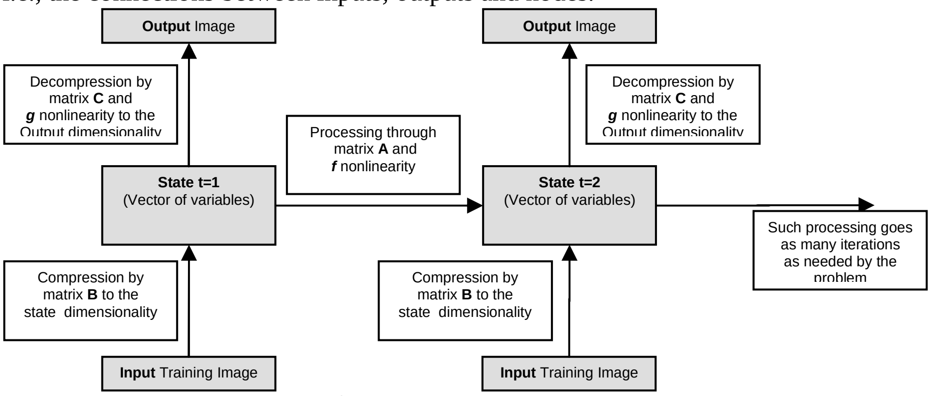


Fig.1. The CVRNN architecture used for pattern recognition. A, B and C are weight matrices. Lines show the connection between the layers. The figure is simplified version of the used architecture to show the connections between the layers. In the experiments the number of states was 50, the figure shows 2 states. White boxes explain the transition,

The CVRNN under consideration can be described in a short form (see eq. 2):

$$\underbrace{\begin{bmatrix} \mathrm{C} & \mathrm{no} & \mathrm{linear} \\ \mathrm{A} & \mathrm{nh} & \mathrm{tanh} \\ \mathrm{B} & \mathrm{ni} & \mathrm{linear} \end{bmatrix}}_{ns} \quad (2)$$

where no is number of outputs, ni is number of inputs, nh is number of neurons in the state and ns is number of states.The weights A, B, C (see Fig. 1 and eq. (1)) should be the same – both in the training as well as in the recognition phase. For this purpose the shared weights concept proposed in [5] was used. Particular care should be taken in the choice of the nonlinearities (functions f, g) since they are unbounded (due to the Liouville theorem; this issue was discussed in [7]). In our work the $\tanh$ function was used for the state interaction and for the state-output interaction $f = g = \tanh\left(\cdot\right)$. The input layer propagates the information as linear function. Matrices *A*, *B* and *C* are sparse matrices with the about 70% zeros. The network was trained using the pattern-by-pattern complex-valued gradient descent training [5]. More information about the complex-valued neural networks as well as about training algorithms and extension of the backpropagation to the complex-valued case can be found in [5 and 7].

3 CVRNN Simulates Network of Oscillating Units

In order to use CVRNNs as models for the binding hypotheses, we coded the images in a “physical” way, which means that each pixel of an image (we have used images 50 by 50 pixels) is now represented by two real-valued numbers which are amplitude

ESANN 2012 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 25-27 April 2012, i6doc.com publ., ISBN 978-2-87419-049-0. Available from http://www.i6doc.com/en/livre/?GCOI=28001100967420.

(strength of the light wave) and phase (color of the light wave) according to the following Table 1.

Table 1. Color coding in the experiment. The modulus is always 0.3. The Phase part varies depending on the color

Color	Red	Green	Blue	Yellow
Coding	$0.3 \times e^{i \cdot \pi/6}$	$0.3 \times e^{i \cdot \pi/3}$	$0.3 \times e^{i \cdot 2\pi/3}$	$0.3 \times e^{i \cdot 5\pi/6}$

After the training is complete, we assume that our network already has some basic ideas about the objects (for example the basic premise about the triangles is a general triangle without the color information) (as inspired by Plato, see [6]). Thus the network should be able to recognize any triangle of any color which comes from the input side.The training was implemented as follows: we used 3 color images, red, green and blue. The shapes used were circles, triangles and rectangles. In order to test how well the network can generalize, we used test images of circle (the shape is from the training set, but not the color) and a yellow rhombus (neither shape nor color are from the training set). The network was trained using the pattern by pattern gradient descent training and error backpropagation algorithm.

Training set: Input and Output Images

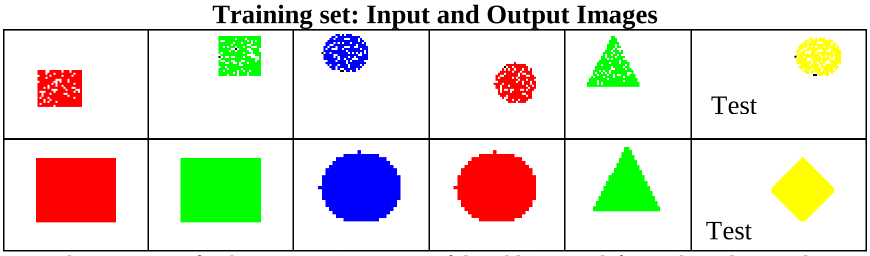


Fig. 2. The training set for the CVRNN (upper row of the table). From left to right: red rectangle, green rectangle, blue circle, red circle, green triangle and yellow circle (we used just a limited set of figures in the paper, the training set contains 15 images of similar type). All images were 20% corrupted by Gaussian noise (while dots at the input images). The background of all images is black. Target output for CVRNN is lower row of the figure. Last image is absent due to generalization reasons. Test images are marked with the word "test". Other upper images are input images, lower are target images.

The experiment consists of two stages: in the first stage, objects are part of the training set and we look at how the color is represented in the network in order to show that neurons of the output layer will be synchronized in phase to represent the color and spatial shape. In the second stage we will look at the network when an object of known shape but unknown color is presented to the network. The task is then to see which neurons will be synchronized and which phase the output neurons will have. Moreover, for all experiments, we track all outputs of the CVRNN in order to see how the memory works in case of image recognition (note that CVRNN has memory due to its structure). The images used for the training in the experiment were static. Clearly, in the general case, objects should be presented to the network from different angles of view and maybe even as a sequence of images. Thus, the network would be able to generalize any triangle to the class of triangles and also to categorize the color by synchronization of the neurons in the triangle class with the color of the object.

4 Results of Perception

The CVRNN (following representation (2)):

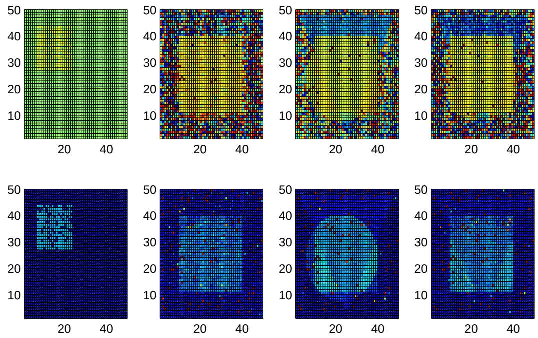
ESANN 2012 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 25-27 April 2012, i6doc.com publ., ISBN 978-2-87419-049-0. Available from http://www.i6doc.com/en/livre/?GCOI=28001100967420.

$$\underbrace{\begin{bmatrix} \mathrm{C} = \left[100 \times 2500\right] & \mathrm{no} = 2500 & \mathrm{linear} \\ \mathrm{A} = \left[100 \times 100\right] & \mathrm{nh} = 100 & \tanh \\ \mathrm{B} = \left[2500 \times 100\right] & \mathrm{ni} = 2500 & \mathrm{linear} \end{bmatrix}}_{\mathrm{ns}=50}$$

was trained for 10 000 epochs using the extended training file similar to Fig. 1 (similar means there were more images with different positions of shapes (circles, triangles and rectangles) in the image and more combinations of RGB colors for the particular shapes). The size of each image was 50×50 pixels. The learning rate was $\eta = 0.02$ .The final training error after the training step (see eq. (2)) was equal to ~ 10^{-11} . All matrices *A*,*B* and *C* were initialized by random weights and were selected to be sparse, analogous to the real world neurons (the sparsity of the biological network is about 20%). In research paper, the sparsity of matrices B and C was 60%, the sparsity of the matrix A was 30%.

4.1 Results of the synchronization of neurons

Next, we will simulate the synchronization of the neurons. The image at the input is a corrupted image of some shape (e.g., the rectangle), which is smaller than the target and is placed at some random place on the input image plane. The output image is a generalized shape of the object (see Fig. 2) of the same color. The training set contains same objects of different colors to make the network's task more complicated. The network is presented with objects which it saw during the training (and with the same color it saw during the training) – but it has never seen both at the same time.
If the input image of a particular color causes the target of the correct shape class to be of the same color as input, then we can say that the network binds the information about the shape and the color. The synchronization simulation results are shown in Fig. 3.



ESANN 2012 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 25-27 April 2012, i6doc.com publ., ISBN 978-2-87419-049-0. Available from http://www.i6doc.com/en/livre/?GCOI=28001100967420.

Fig. 3. Example of neuron synchronization for the object from the training set. Upper row left to right: argument of the input pattern, argument of the network output from the 3rd state, phase of the network output from the last 50th state. Lower row, left to right: modulus of the input pattern, modulus of the output from the 3rd state, modulus of the output from the 50th state. The bar on the right shows the color ranges for all pictures. We see, that input phase is equal to the network output phase, which implies binding of color and shape.

One can see from Fig.3 that neurons synchronize not only in phase, but also in the spatial distribution representing the color and the shape of the object. Therefore, there is a coupling of the phase and the spatial distribution. This means that a particular input from the training set can cause the correct output, which corresponds to a generalized shape (see Fig. (1)) with the color of the input object. In other words: CVRNN offers a nice and elegant solution to the binding problem.

4.2 Capability of the network to recognize known object with unknown color

As shown, neurons synchronize for the same color. Moreover, the spatial structure of the synchronized neurons represents the shape of the target object. Now, we can proceed and try to force the network to generalize objects, which means present an image to the network, which it has not seen during the training and observe whether it will be able to recognize the shape and the color. The results are shown in Fig. 4. The network input is the yellow noisy circle. One can see from Fig. 4 that the network was still able to recognize the shape, but not the color (further experiments have been conducted for different shapes which are not in the training set). This can be explained as follows: the network can generalize, but without the color, if it has never seen the color before (see Fig.4). Since it would be beyond the scope of this paper, we cannot show all the training sets and results.

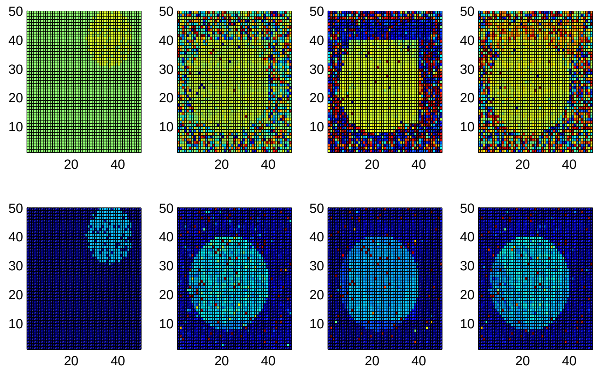


Fig.4. Example of the neurons synchronization for the pattern from the training set. Upper row left to right: input pattern – argument, network output from the 3rd state – argument, network output from the last 50th state – argument. Lower row, left to right: modulus of the input pattern, modulus of the output from the 3rd state, modulus of the output from the 50th state.

ESANN 2012 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 25-27 April 2012, i6doc.com publ., ISBN 978-2-87419-049-0. Available from http://www.i6doc.com/en/livre/?GCOI=28001100967420.

4.3 Capability of network to recognize unknown objects with unknown colors

Now we consider what happens when an unknown object is paired with an unknown color and presented to the network: the input object is a yellow rhombus. The simulation result shows (here we do not represent the figure due to the paper limitations) that the network produces a noisy output, which consists of two shapes, namely the circle (with maximum possible phase) and the triangle, with minimum possible phase. The modulus of the output is very noisy with some shadows of circle and triangle from the training set. The phase picture is easier to recognize. The interpretation of the two shapes seen in figure is that the network tries to remember the most topologically similar objects from the memory which exists due to the states interaction. The triangle and the circle appear due to the fact that the structures in the data matrix containing the triangle or the circle are very similar to the rhombus (here, we use a low resolution, which is 50 by 50). Therefore an overlay of both shapes appears as output. If the rhombus was used for the training of the network, it can be reproduced with its color.

5 Summary and Outlook

We have shown that CVRNN can simulate the synchronization of weights combining the information about the shape of the object together with information about its phase. Moreover, the memory, the generalization and the synchronization can be simulated by the interaction of the phase of complex values by using the CVRNN.

Future work should include/focus on online training of the CVRNN in order to use the information about the object by looking at it from various perspectives. This will enableus to create a very large data base of images, and we will be able to expand the generalization capabilities of the CVRNN for more classes of objects.

References

[1] C. v. d. Malsburg. The What and Why of Binding: The Modeler's Perspective. Neuron, Vol. 24, 95–104, September, 1999

[2] Maye A. Correlated neuronal activity can represent multiple binding solutions. Neurocomputing, 52-54:73-77, 2003.

[3] L. Ward. A. K. Engel, P. R. Roelfsema, P. Fries, M. Brecht and W. Singer. Role of the temporal domain for response selection and perceptual binding. Cereb. Cortex (1997) 7 (6): 571-582. doi: 10.1093/ cercor/7.6.571

[4] Synchronous neural oscillations and cognitive processes. Trends Cognitive Science. 2003 Dec; 7(12):553-9

[5] G. Zimmermann, A. Minin., V. Kusherbaeva., Historical Consistent Complex Valued Recurrent Neural Network, ICANN 20011, Part 1, LNCS 6791, pages 185-192, 2011.

[6] L. Perlovsky, Neural Networks and Intellect, Using Model Based Concepts, 469 pages, Oxford University Press, 2001,

[7] Zimmermann H.-G., Minin A., Kusherbaeva V., Comparison of the complex valued and real valued neural networks trained with gradient descent and random search algorithms, Proc. of ESANN 2011, pp. 216-222, 2011.