Neural Architecture for Complex Scene Recognition Based on Rank-order Features of IT Neurons

Sergey Tarasenko and Natalia Efremova

Abstract—Human brain is an information processing system, which is perfectly designed to deal with complex visual scenes. We propose a novel architecture for object and place recognition, taking inspiration from the primate ventral visual stream (areas VI-IT). The functionality of the system is based entirely on recent neurophysiological findings and is implemented by means of biologically plausible information processing mechanisms. We illustrate the ability of the system to recognise multiple objects within various positions in the retinal image. During the experiments, we show that the network can learn to recognise the position of object, in which it appears most frequently. Simulation results are consistent with the animal experiments. The above-mentioned properties of the network demonstrate classification and rank-object preserving properties of the neurons in the IT region.

I. INTRODUCTION

HUMAN brain is capable of processing stimuli of various modalities both separately and fused together. The brilliant example of the prior one is visual information processing in ventral and dorsal streams [1], [2], [3], while mirror-neuron system [4], [5] can serve as an example, when stimuli of several modalities (motor, visual, sensory) are processed simultaneously.

Nevertheless, the recognition and classification tasks are preformed with extreme accuracy. Recent neurophysiological studies have provided solid core of evidence that visual information processing is also very rapid. It takes from 100 to 150 ms for humans, and this time is even shorter for higher primates [6], [7].

Brain’s super-accurate performance and super-short processing time are the two major reasons, which make brain-inspired information processing architecture a popular tool for visual information processing. A great variety of studies have been concentrated on modelling of particular features of the visual information processing in the primate brain [8]-[9].

On the other hand, human brain mechanisms provide opportunities for both global formation of the brain’s topology across various brain areas[10], [11] and local fine tuning of the particular connection within a given brain area [12]. The notions of local and global information can differ regarding the context. Furthermore, this segregation between information types depends on the goal of simulation and the precision of analysis. In this paper, we define the term “local information” as the properties of particular object or parts of objects, whereas “global information” describes mutual allocations of all the objects throughout environment.

The ability of the information processing system to vary the number of processing units in time is essential for dealing with complex scenes. It has been shown, for example, by Platt [13] and Fritzke [14] that performance of neural networks with variable number of neurons is more precise than the performance of the networks with fixed number of neurons when dealing with non-stationary inputs. Such supremacy in performance is explained through the ability to encode various distant areas in the feature space by including new processing units.

To sum up, we conclude that there are three key features to be incorporated by the information processing system in order to deal effectively with changing environment. These are 1) rapid signal propagation and processing; 2) super-accurate processing, and 3) ability to adaptively change its own internal structure.

In this study we only concentrate on the visual information processing. Therefore, hereafter the notion of environment is equivalent to the notion of Visual Scene (VS). In the previous studies, we have presented a general framework for development of the cortex-like visual object recognition systems [15]. The framework incorporates the information processing principles, which are recognized to be intrinsic for the primate brain. The proposed framework is based on the functional architecture of the Visual Ventral Stream (VVS) of the human brain.

However, from the point of view of local vs global information processing, this general framework is only capable of dealing with individual objects. In other words, system can only deal with local information by identifying different objects, but cannot learn and recognize various properties of objects’ appearance in environment. Therefore, our objective is to extend the existing framework in order to provide an additional functionality for processing of objects’ global information. We intend to introduce additional processing modules, which will implement detection and recognition of objects’ mutual allocations in the visual field.

The major basis for improvement of the existing framework stems from the recent results of neurophysiological studies on Inferior Temporal cortex (IT) [16]-[17].

II. BACKGROUND

This paper regards some basic principles of object recognition, which are used in primate visual cortex. In order to outline these principles, we will first briefly describe the region of interest.
In primate cortex, the visual information is processed in two major parallel processing streams: ventral and dorsal pathways [18]. The ventral pathway (VVS) is considered to be responsible for “object vision”, i.e., discrimination and recognition of visual images of objects [1]. In primate cortex, it is presented by the number of areas, crucial for visual perceptual processing and recognition of objects: V1, V2, V4, and IT, organized in a retinotopic manner [19].

The retinal image is first processed by primary visual cortex (V1). V1 neurons are selective for a number of features, the most important of which is orientation [20]. Therefore, we can model this area with Gabor-like filters, which are reported to be similar to the receptive fields of primate cortex [21]. V1 sends feedforward signals to many higher visual areas such as V2, V3 and V4. Shape representation in area V4 is distributed between individual neurons, encoding smaller parts of larger objects [22]. Kobatake and Tanaka quantitatively examined visual responses in areas V2, V4, anterior IT and posterior IT [2]: the results of the study indicate that, first, these areas are activated by presentation of stimuli of different complexity, and next, the sizes of the receptive fields (RFs) in these areas differ dramatically.

We will employ these findings to outline the relative roles of cortical areas V1-V4 in our model.

It is widely accepted that neurons in IT can be activated by the presentation of isolated visual objects in different parts of the retinal position. Recently, the large number of psychophysical studies were devoted to understanding of the response properties of IT neurons during the presentation of multiple objects. Neurons in the IT area possess the ability to preserve each neuron’s rank-order object selectivity across object position changes and clutter conditions [16]. It was shown, that neurons in the TE area maintain their rank-order object selectivity within their receptive field (RFs) irrespectively of the changes in object position [23], [24]. In this paper, we aimed to carry an extensive simulations, based on the results of the above studies.

The functionality of the proposed network is based entirely on these recent neurophysiological findings and is implemented by means of biologically plausible information processing mechanisms.

III. SYSTEM ARCHITECTURE

Our system consists of four processing levels (modules). Three of these levels are inherited from the framework presented in [15]. Therefore, the details about the information processing in these modules will be mentioned only briefly. Original framework consists of three levels, which correspond to V1, V4 and IT areas of the VVS. In the present system, processing modules V1 and V4 have remained unchanged, while the functional concept and architecture of IT module has been essentially reconsidered. The IT module in our recent system includes two sub-modules, which correspond to posterior (TEO) and anterior (TE) regions of Inferior Temporal cortex. The entire architecture of the system considered in this paper is presented in Fig. 1, and the additional part is highlighted by the area ”Additional module”.

The details of processing in new modules, intra- (TE) and inter- (between TEO and TE) module interactions are discussed in subsequent sections. The details of interaction of inherited modules are presented in [15] and briefly outlined by legend of Fig.1.

Fig. 1. Detailed scheme of system architecture. Brief explanation of signal propagation through the system: the V1 module consists of four distinct Gabor filter maps, which are mutually inhibited (round-head solid lines). After inhibition the outputs of all four maps are integrated by the sub-module of the V4 module. The output of this sub-module is Integrated Orientation Map (IOM). The IOM is then sent to the V4 growing SOM sub-module by means of excitatory connections, which are indicated by arrow-head solid lines throughout the scheme. Next, IOM is translated into the Feature Map (FM). FM is sent to V4RBF sub-module and TEO growing SOM (TEOSOM) sub-module. The TEOsub-module receives excitatory inputs from the TEO growing SOM units as well. The TEOsub sends modulating (amplification of existing activation level) signals (marked with diamond-head dashed lines) back to the V4 RBF units. The feedback modulated (amplified) signal from V4 RBF units to TEOsub units is marked with arrow-head double solid line. The V4 and TEO growing SOM units serve as repositories of visual features of objects. Interaction between the V4 and TEO growing SOM units updates objects with valid features (indicated by double-end arrow-head dashed line). TEO RBF units send excitatory signals to TE units. The signals from TE units are then propagated into the “recognise object and detect location” module.

The functionality of the proposed network is based entirely on these recent neurophysiological findings and is implemented by means of biologically plausible information processing mechanisms.

A. Information processing in TEO module

The TEO module contains TEOSOM and TEOsub modules. Each TEOSOM neuron contains the maps of distinct objects. Therefore, number of TEOSOM neurons is equal to number of the stored objects: TEOSOM(i), where i is a number of objects, i = 1, ..., n, and n is a total number of objects stored in the system.

The TEOsub neurons are used to compare the actual visual input and the objects, stored in the system. There are n different classes of TEOsub neurons: TEOsub(i), where i = 1, ..., n. Each class corresponds to a single object. The response of TEOsub neuron is calculated as:

\[ TEOsub(i) = \exp(-\beta_{TEO}|X - TEOSOM(i)|^2) \]  (1)

where X is an actual input into a TEOsub neuron, and \( \beta_{TEO} \) is a scale parameter. The parameter \( \beta_{TEO} \)
is calculated by using Feature Map (FM) as: \( \beta_{\text{TEO}} = (\sum_{k=1}^{m} \sum_{p=1}^{n} \Theta(F_{\text{eel},ij}))^{1/2} \), where \( \Theta(\cdot) \) is Heaviside function, and \( F_{\text{eel},kp} \) is a feature number located at intersection of \( k \)-th row and \( p \)-th column of FM matrix (see [15]).

The \( \text{TEO}_{\text{RBF}} \) sub-module takes its input from the V4 module and \( \text{TEO}_{\text{SOM}} \) sub-module. The V4 module sends \( \text{FM} \), which is a table containing numbers of visual features that fits the best input visual signal at given location (see [15]) to the \( \text{TEO}_{\text{RBF}} \) module. FM is sampled by the RFs of \( \text{TEO}_{\text{RBF}} \) neurons. Therefore, actual input \( X \) is a part of a FM.

The \( \text{TEO}_{\text{RBF}} \) sub-module produces a collection of response grids. A single response grid illustrates activations of \( \text{TEO}_{\text{RBF}}(i) \) neurons of a certain class at each spatial location of FM.

In order to identify the object in VS the following steps are required: 1) select maximum activation value in a response grid \( i \) and assign it to variable \( maX_i \), where \( i \) is a number of a class; 2) select maximum value from \( maX_i \) and assign it variable \( maX: maX = \max\{maX_i\} \).

If \( maX \) value is greater then the threshold \(^1\), we identify the number of the object: \( i^* = \arg\{maX\} \).

If \( maX \) value is smaller than the threshold, then a new object is stored in a new \( \text{TEO}_{\text{SOM}}(n + 1) \) neuron.

The detailed explanation of TEO module function is presented in [15], where TEO module is referred to as IT module.

**B. Information processing in TEO**

We define two types of the TE neurons. The neurons of the first type produce the overall output of the TE module, while neurons of the second type perform auxiliary function. The neurons of the first type are called sigmoid TE neurons and denoted as \( \text{TE}_{\text{sig}} \). The neurons of the second type are inhibitory interneurons and are denoted as \( \text{TE}_{\text{inh}} \). Each \( \text{TE}_{\text{inh}} \) is bound to a particular \( \text{TE}_{\text{sig}} \) Neuron, therefore we have the equal number of neurons of each types. The \( \text{TE}_{\text{sig}} \) are used for place-object recognition and perform rank-ordering, while the inhibitory neurons are used to inhibit the activation of the sigmoid neurons. The signal propagation in the TE module is based on mechanisms of Hebbian learning and shunting inhibition (see below).

Here we present a scheme of connections between the \( \text{TEO}_{\text{RBF}} \) neurons and the \( \text{TE}_{\text{sig}} \) neurons. Since there are several types of \( \text{TEO}_{\text{RBF}} \), single TE neuron is connected to the response grids of \( \text{TEO}_{\text{RBF}} \) neurons. However, only values \( maX_i \) are propagated to the \( \text{TE}_{\text{sig}} \) neuron. Therefore, a single TE neuron takes only \( n \) inputs from \( \text{TEO}_{\text{RBF}} \) sub-module. A schematic interaction between a single pair of \( \text{TE}_{\text{inh}} \) and \( \text{TE}_{\text{sig}} \) neurons with \( \text{TEO}_{\text{RBF}} \) neurons of each class is depicted in Fig. 2.

In contrast to \( \text{TEO}_{\text{RBF}} \) neurons, which densely cover the VS with RFs, RF of distinct \( \text{TE}_{\text{sig}} \) and distinct \( \text{TE}_{\text{inh}} \) neurons do not intersect. On the other hand, neurons \( \text{TE}_{\text{sig}}(j) \) and \( \text{TE}_{\text{inh}}(j) \), have the same RFs (\( j = 1, \ldots, m \), and \( m \) is a total number of \( \text{TE}_{\text{sig}} \) and \( \text{TE}_{\text{inh}} \) neurons). Therefore, each pair of \( \text{TE}_{\text{sig}} \) and \( \text{TE}_{\text{inh}} \) neurons takes input from a particular location of VS.

The activation of a single \( \text{TE}_{\text{sig}} \) neuron is modelled by sigmoid function:

\[
\text{TE}_{\text{sig}} = \frac{1}{1 + e^{-(\text{Input} - w)w}}
\]

where \( w \) is a weight vector \( w = \{w_1, \ldots, w_n\} \), and \( \text{Input} \) is an input vector \( \text{Input} = \{maX_1, \ldots, maX_n\} \).

The RFs of TE neurons used in our simulation are shown in Fig. 3.

**Hebbian learning.** Hebb has suggested that certain cortical cell populations behave as functional units with coordinated activity patterns in accordance with changing synaptic strength [26]. It was assumed that the strength of a synapse (connection) between neuron 1 and neuron 2 increases, when firing in neuron 1 is followed by firing in neuron 2 with a very small time delay.

The discrete Hebbian learning rule for a single neuron can be presented in the form of eq. (3):

\[
\Delta w(k + 1) = \mu[y(k)x(k) - \alpha w(k)]
\]
where \( y \) is output of the neuron, \( w = (w_1, \ldots, w_n) \) is vector of synaptic weights, \( \mu \) and \( \alpha \) are constants regulating learning and forgetting rates, respectively, and \( k \) indicates the iteration number. The output \( y \) can be calculated as \( y = f(s) \), where \( f \) is activation function, \( s = w^T x \), and \( w = (w_1, \ldots, w_n) \) is a vector of synaptic weights, \( x = (x_1, \ldots, x_n) \) is vector of inputs, and \( n \) is a total number of inputs of the given neuron.

The synaptic weights are updated according to the following equation:

\[
w(k+1) = (1 - \alpha \mu)w(k) + \mu y(k)x(k) \tag{4}
\]

In all subsequent simulations we set parameters of learning and forgetting rates as follows: \( \mu = 0.3 \) and \( \alpha = 0 \).

**Shunting Inhibition.** Our goal is to uniquely identify a single object at the particular location of the VS. However, the realization of pure Hebbian learning is not enough for that: the activation of the neuron is based on the weighted sum of all inputs. Therefore, the output of the neuron is based on the information about all the objects, and such approach is not sufficient to identify a single object.

To solve this problem, we suggest to use rank order approach, which implies that the neuron response time depends on the magnitude of the its output. In other words, it takes shorter time to produce a stronger response than a weaker one. If responses of the neurons on the previous processing level are different, then the signals from the neurons with stronger responses should reach the neurons on the next level faster than the signals from neurons with weaker responses [6].

To take an advantage of this time difference in order to uniquely identify a single neuron with the strongest response, Thorpe [27] has suggested to use shunting inhibition. The inhibitory interneurons can be used for modulating of the activity inside the neuron. Thorpe has suggested that rank order approach and shunting inhibition taken together can solve the “winner-takes-all” problem. According to Thorpe, the first signal to arrive is not only propagated into a neuron, but also directed into inhibitory interneuron. The inhibitory interneuron sends its projection to the neuron, therefore, the second signal to arrive is inhibited by the first signal, and so on. Therefore the first signal to arrive is the only one that stays unaffected.

We intend to adopt the mechanism of shunting inhibition in our model. To do so, we modify the mechanism of calculating response of \( TEO_{sig} \) neurons. First of all, the vector of inputs \( x \) contains the signals with different arrival time. We consider that the sequence of components of vector \( x \) corresponds to the order of signal arrival, i.e., the 1st component corresponds to the first signal to arrive, the 2nd component corresponds to the second signal to arrive, etc.

We consider that the weight associated with connections between an inhibitory interneuron \( TEO_{inh} \) and \( TEO_{RBF} \) neurons equals 1 and is fixed. Therefore, the input signal is propagated into interneuron without any change. Connection weights between \( TEO_{inh} \) and \( TEO_{sig} \) are also set to 1.

The inhibition is formalized as subtraction of one signal from another. However, the signal arriving earlier has bigger magnitude, therefore the inhibition will always result in negative value. Without loss of generality, we consider that result of inhibition is non-negative. Therefore, we multiply a result of each inhibition by output of Heaviside function \( (9) \).

Considering the above explanation, variable \( s \) should be redefined as:

\[
s = \sum_{i=1}^{n} w_i(x_i - x_{i-1})\Theta(x_i - x_{i-1}) \tag{5}
\]

where \( n \) is a number of objects in the system \( (n = 8) \) and \( x_0 = 0 \). The mechanism of Hebbian learning with shunting inhibition allows simulation of development of connection between \( TEO_{RBF} \) and \( TEO_{sig} \) neurons. At this level of the framework, development is considered as dynamical process of changing connection weights between TE neurons and their inputs (afferents).

Here we discuss in details the evolution of weights between a single \( TEO_{sig} \) neuron and \( TEO_{RBF} \) neurons. One \( TEO_{sig} \) neuron receives inputs from one inhibitory interneuron \( TEO_{inh} \) and eight \( TEO_{RBF} \) neurons with maximum activations. We set the initial weights between \( TEO_{RBF} \) neurons and \( TEO_{inh} \) neuron to be all equal to value 1.

To prove ability of the framework to identify correctly the object in both TEO and TE modules, we provide simple quantitative result. If the input from \( TEO_{sig} \) equal 1, then \( x_1 = 1 \). In such case all the subsequent inputs will be inhibited by the unit signal \( x_1 = 1 \), and the output of the \( TEO_{sig} \) neurons is calculated as \( \frac{1}{1 + exp(-1)} = \frac{1}{1 + exp(1)} \approx 0.73 \).

Thus, output of \( TEO_{sig} \) neurons is greater than the threshold 0.68. Therefore system can correctly identify object in both TEO and TE modules.

**IV. Simulations**

**A. Stimuli**

In the following simulation experiments, we used 2D projection of computer-generated 3D objects (Fig. 4 a). These stimuli were presented to the network one by one, so the system could learn each object separately. Then, we have shown that original system could recognize multiple objects anywhere in VS. VS contained five different objects, four of which were presented only once, while the fifths object was presented twice. The response grids of corresponding \( TEO_{RBF} \) neurons and VS are presented in Fig. 5. The activation of value 1 (greater than threshold 0.68) indicates correct recognition of an object at a certain location of the VS. Such behaviour could be achieved also with the original framework.

**B. Experiment 1: modelling place-object selectivity of \( TEO_{sig} \) neurons to column stimuli**

According to the recent research, the TE neurons exhibit the place and object selectivity [16]. This means that the response of the neurons depends on object type and spatial location. Li and colleagues [16] have proposed several simplified SMV classifiers, which model the phenomenon.
The peculiarity of the classifiers is that they can grasp the frequency of appearance of a particular object at a particular spatial location of the VS.

In this study, we are going to illustrate that recently considered $TE_{sig}$ neurons are capable of producing adaptive response bound to object location and type. In our model, adaptiveness of responses is emergent property of a neural population. We apply Hebbian learning with shunting inhibition to show how such property of a single $TE_{sig}$ neuron can emerge on the basis of biologically plausible interaction within surrounding neural population.

First experiment was designed to show the network’s ability to recognize objects, presented at a specific location in VS. In the simulations, we tried to mimic the experiment, described in [16]. For these purpose, we used slightly modified object layout and experimental procedures (Fig.6).

For the purposes of experiment 1, we have designed special stimuli, each of which was composed of three cells, organised in a column (Fig. 4 c). A single cell corresponds to RF of $TE_{sig}$ neurons. A size of each cell is 100-by-100 projections from $TEO_{RBF}$ neurons. The objects in the cells were selected randomly. In total we have created 150 column stimuli. These stimuli were presented one by one to the feed-forward network, which simulates interaction between $TEO_{RBF}$ and $TE_{sig}$ submodules. The result of this simulation is the distribution of weights between the $TE_{sig}$ and $TEO_{RBF}$ neurons. Presentation of a single stimulus corresponds to a single epoch of network learning.

The resulting weights are presented in Fig. 7. The default response of the all $TE_{sig}$ neurons before simulation is equal to 0.73 (see description of shunting inhibition). Fig. 8 depicts the responses of $TE_{sig}$ before and after learning.

We have discussed above that identifications of object by both $TEO_{RBF}$ and $TE_{sig}$ neurons are equivalent in terms of correct recognition of the same object. On the other hand, neurons in $TEO_{RBF}$ are tuned to recognize only a single object regardless of its location and frequency of appearance. In contrast, $TE_{sig}$ neurons have multiple inputs from $TEO_{RBF}$ neurons.

Hebbian learning with shunting inhibition allows to monitor and map the frequency of appearance of each object in the RF field of a given $TE_{sig}$ neuron. $TE_{sig}$ neuron themselves are bound to a particular location in the VS. Therefore, the distribution of weight values between the given $TE_{sig}$ neuron and the $TEO_{RBF}$ neurons allows to code information about frequency of appearance of a certain object at a given location.

On the other hand, strength of connection between certain $TEO_{RBF}$ and $TE_{sig}$ neurons demonstrates the selectivity of the latter to a particular object: the higher is the weight the higher is selectivity to a particular object.

The recognition by $TEO_{RBF}$ and $TE_{sig}$ neurons is in coherence, meaning that the same object is recognized correctly. But in contrast to the constant response of $TEO_{RBF}$ neurons, responses of $TE_{sig}$ neurons vary (Fig. 8) due to adaptive mechanism described in this paper. Therefore, we consider TE neurons as modulators or gain functions for responses of $TEO_{RBF}$ neurons.

Therefore, the feed-forward network between $TEO$ and TE sub-modules allows modelling selectivity of the $TE_{sig}$ neurons to objects by means of encoding frequency of object appearance by weight values. Since the $TE_{sig}$ neurons are bound to a particular retinal position, they code the spatial location of the object as well. These two pieces of information together allow to encode rank-order selectivity of the neurons across object positions.

Therefore the results of this experiment demonstrated: 1) the ability of the network to recognize the objects in a complex scene, irrespectively of presence or absence of other objects in the VS, and 2) the ability to learn the place of occurrence of the objects in the VS.

Finally, we demonstrate ability of our system to answer the question, which are similar to the ones used in study [16]. The first question was “Is a particular object present in any position?” (Fig. 6, b). This question is answered positively in Fig. 5.

The second question was “Is a particular object present in a particular position?” In Fig. 6, c, we show sample stimuli which provide positive and negative answers to the question “Is the star present in the top position?”.
Until now, we only considered the mechanisms behind the phenomenon of the rank-order selectivity. We have shown that our network can successfully demonstrate such selectivity under conditions of neurophysiological experiment. Next, we illustrate how to use our network in order to answer particular questions about object type and object position.

In Fig. 9, we illustrate a procedure, which can be used to get the answer to the question “Is the star present in the top position?”. This question can be reformulated in the form of conjunction of two questions: “Is an object present in the top position?” AND “Is an object a star?”.

In our system, the answer to the second question is given by the TEO_RBF neurons. First, we obtain response grids of all TEO_RBF neurons within a RF of a given TE_sig neuron. Then we take maximum values $\text{max}_i$, $i = 1, ..., 8$ for all response grids. Next, we calculate value $\text{max}$ as a maximum among values $\text{max}_i$, $i = 1, ..., 8$. Finally, we identify the object number $i^*$, which corresponds to the value $\text{max}$.

The answer to the first question is given by TE_sig neurons. In this example, we use three TE_sig neurons. Their RFs are located at top, middle and bottom positions, receptively. Therefore, we need to check whether response of a TE_sig neuron, whose RF corresponds to the top position, is greater than a threshold value.

In our example, the value of weight between TEO_RBF neurons tuned to recognize star object and TE_sig neuron in the top position is 6.89 after learning. Since, value $\text{max}$ is 1, then response of TE_sig in the top position is calculated as: $1 + \frac{1}{1 + \exp(-6.89 \times 1.00)} \approx 0.999$. This value is greater than a threshold. Therefore, TE_sig confirms recognition of an object in the top position.

Therefore, answers to both questions “Is an object present in the top position?” AND “Is an object a star?” are positive. Consequently, our system can give an answer “Yes” to the original question “Is the star present in the top position?”, which is correct answer.

To sum up the experimental results, we conclude that performance of our new network complies with neurophysiological results presented in [16]. Therefore, we could successfully achieve rank-order selectivity by using biologically plausible processing algorithms.
The task of detection of an object at a particular location is accomplished by means of two types of neurons. Neurons of \( TEO_{RBF} \) are used to identify the object itself. In this figure, the maximum activation of \( TEO_{RBF} (3) \) indicates that the object is a star. Neurons of \( TE_{sig} \) type are bound to a particular position in a VS and used to identify a particular position of the object, which was “recognized” by \( TEO_{RBF} \) neurons. In this case, the \( TE_{sig} \), which corresponds to the top position, is the only neuron with activation 0.999 greater than threshold value. Therefore, top position is chosen as the location of the “recognized” star (object 3).

We illustrated the ability of our system not only to recognize eight types of stimuli within any position in the VS, but also to preserve the rank-order selectivity of \( IT_{sig} \) neurons, which has been reported in recent neurophysiological studies. We have shown that the network learned not only to recognize the objects, but also to remember the place in a VS, in which the object appeared, and the frequency of its appearance.

The model was trained to perform several recognition tasks that were considered to demonstrate the innate properties of the visual cortex. It shows, that a) our model is neurophysiologically plausible, and b) the properties of IT neurons, which are widely accepted to be intrinsic to the primate brain, can be modelled with the proposed architecture.

V. DISCUSSION

In this study we extended a general framework for development of cortex-like visual object recognition system. We illustrated the ability of the new system to recognize different objects in various positions in the VS. During the experiments, we also showed that the network could learn to recognize the position of objects. The object recognition at a particular location is based on the frequency of the object network.
appearance at this location. We have proposed a biologically plausible learning algorithm based on Hebbian learning and shunting inhibition. This algorithm allowed to map the frequency of object appearance in a particular position onto the space of synaptic weights between $TEORBF$ and $TEISIG$ neurons. Therefore, we have successfully incorporated the frequency information into the dynamics of synaptic weights evolution over the time.

Moreover, we showed that the addition of the new modules to the proposed network augments the ability of the system to analyse complex scenes in the VS. This is achieved with the combination of the mechanisms of Hebbian learning with shunting inhibition, which allows to simulate evolution of the synaptic between $TEORBF$ and $TEISIG$ neurons [27].

The proposed network exhibited the ability to recognise different objects in various positions in the VS. These results are consistent with the animal experiments [16], [28]. Therefore, the network demonstrated classification and rank-object preserving properties of the neurons in the TE region.

These properties have been already modelled with various types of network architectures [29], [30]. We showed that such task can be accomplished in a biologically plausible way by means of the proposed architecture as well. A number of studies are dedicated to recognition of complex VS: a good example of such a study is by done Rolls and Deco, in which the authors show how the inferior temporal visual cortex operates to enable the selection of an object in a complex natural scene [31]. The main difference of our system from the above architectures is the dynamic dictionary of features and the ability to learn new objects, while preserving the rank-object selectivity of the neurons.

Finally, it is important to note, that our simulations are based on randomly generated visual scenes. Therefore, our system demonstrated the capability of extracting the regularities from stochastic visual experience.

Summarizing, we can conclude that our architecture allows processing of both the local and the global information in the VS. The local information processing implies recognition of the particular object or parts of objects at the VS, whereas global information processing refers to the understanding of mutual allocations of all the objects throughout VS.

REFERENCES