2014 International Joint Conference on Neural Networks (IJCNN)
July 6-11, 2014, Beijing, China

WWN-9: Cross-Domain Synaptic Maintenance and Its Application to Object Groups Recognition

Qian Guo, Xiaofeng Wu, and Juyang Weng

***Abstract*—Where What Network 6 (WWN-6) has shown that its model of synaptic maintenance using neural transmitters acetylcholine (ACh) and norepinephrine (NE) enables each neuron to distinguish between neuronal input lines from its relatively stable object patch and those from irrelevant backgrounds. However, it is about only a single domain — sensory domain X. During development from conception through fetus and newborn, every brain neuron has three major domains of input, sensory X, lateral Y and motor Z. The single-domain model of WWN-6 is not directly applicable to multiple domains because different domains have very different dimension and signal variations that cannot be directly compared. We believe that cross-domain synaptic maintenance is a crucial mechanism to develop a shallow-and-deep processing hierarchy in the brain where each neuron autonomously select domains in the developing hierarchy, not necessarily directly connected to receptors in X and muscles in Z. In the new work here, we propose a biologically inspired model for cross-domain synaptic maintenance. We assume that the earlier connection guided by morphogen result in initial coarse connection, but cross-domain synaptic maintenance refine connections to enable each neuron to autonomously find its role. As concept patterns emerge in Z, neurons refine their connections, to differentiate their roles among sensory processing, motor processing, and a mixture of both. Experimentally, we show the effect of the new theory through learning of individual objects and object groups, where neurons initialized for object-group connections tend to find their receptor inputs from X are not as stable as inputs from motor Z, thus, gradually turn into "later" processing neurons — for "higher-level" object-based features and their invariances. In principle, WWN-9 tends to learn a new object group without repeating the learning of all instances of each individual object.**

I. INTRODUCTION

The research on autonomous mental development aims at studying the developmental mechanisms, architectures and constraints that allow lifelong and open-ended learning of new skills and new knowledge in intelligent agent. As in human children, learning is expected to be cumulative and of progressively increasing complexity, and to result from self-exploration of the world in combination with social interaction. As the autonomous mental development is a sub-field of Artificial Intelligence (AI) which is inspired by human intelligence, more and more models proposed imitate the brain to different degrees. General purpose object recognition and attention in complex backgrounds is one of the significant issues in the field of AI. Since human can accomplish such tasks easily, the model inspired by human vision system is thought as one possible approach to address this open yet important vision problem.

With the advances of the studies on visual cortex in physiology and neuroscience, several biological-inspired network models have been proposed. Applied some biological mechanisms such as lateral competition (winner-take-all), edge features extraction by Gabor filters like what simple cells do, feature combination like what complex cells do, the feed-forward network HMAX by Tomaso Poggio [4] simulates the ventral pathway in primate vision system to do object recognition with the loss of object location. The deep-learning network by Hinton and coworkers [2] adopted the hierarchy in both architecture (multiply hidden layers) and processing flow (deep learning algorithm) in human brain to acquire much better results than traditional shallow networks. But this model is difficult to utilize top-down attention and do many confirmations empirically rather than theoretically in learning. ART by Grossberg and coworkers [1] provided a solution of the stability-plasticity dilemma; namely, how a brain or machine can learn quickly about new objects and events without just as quickly being forced to forget previously learned, but still useful, memories. But it is too sensitive to the change of input.

Different from all above models, WWNs introduced by Juyang Weng [6] is a biologically plausible developmental model designed to integrate the object recognition and attention namely, what and where information in the ventral stream and dorsal stream respectively. It uses both feed-forward (bottom-up) and feedback (top-down) connections. With the lateral inhibitions (competition), the networks can sort out the best-match so that only the near memory is modified and other larger memory is intact as the long term memory for this input context. Moreover, multiple concepts (e.g., type, location, scale) can be learned concurrently in such a single network through autonomous development. That is to say, the feature representation and classification are highly integrated in a single network.

In order to reduce the interference from leaked-in background pixels, the mechanism of synaptic maintenance [5] was proposed in WWN-6 to automatically determine and adapt the receptive field of a neuron. The network intends to retain a subset of synapses that provide a better majority of matches and cut other synapses. However, the above work

Qian Guo and Xiaofeng Wu (Corresponding author), are with Department of Electronic Engineering, Fudan University, Shanghai, 200433, China, (email: {13210720028, xiaofengwu} @fudan.edu.cn); Juyang Weng is a visiting professor at the School of Computer Science, Fudan University, Shanghai, 200433, China and a professor at Department of Computer Science and Engineering, Cognitive Science Program, and Neuroscience Program, Michigan State University, East lansing, Michigan, 48824, USA, (email:weng@cse.msu.edu); This work was supported by Fund of State Key Lab. of ASIC & System (11M-S008) and the Fundamental Research Funds for the Central Universities to XW, Changjiang Scholar Fund of China to JW.

Fig. 1. The motivational pictures of object groups

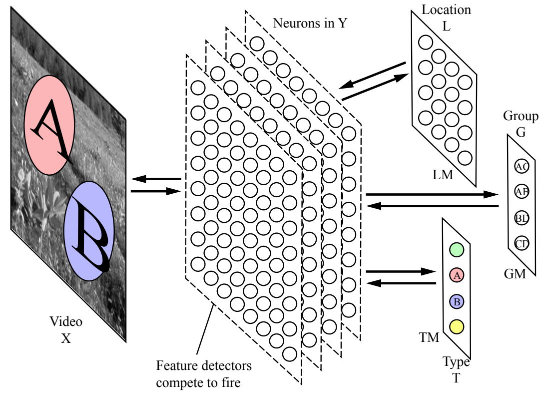


Fig. 2. The architecture of WWN-9 at birth

only involves synaptic maintenance in the bottom-up input from X area.

In the other hands, there are huge numbers of concepts which are correlated with each other in our mind as we learn and interact with the external world. Some concepts are the intuitive reflection of objects while some concepts which are abstracted or summarized by a set of relatively simple concepts usually have no direct connection with visual inputs. Therefore, different concepts in motor area are correlative to some extent, not fully independent. For example, a complex definition in mathematics is based on simple definitions. We teach a child to do the addition by counting fingers where the combination of fingers is the inputs of vision. When it comes to multiplication, we will say "5 multiply 2 equals doing the addition of 2 5 times". In this case, the definition of multiplication based on addition has nothing to do with the bottom up input from vision.

Base on such considerations, we believe that for the network, neurons in Y have synapses bidirectionally connecting with both X and Z at the time of birth, which can combine information from bottom-up input and top-down input. With the further exploring of external world, some of them may only use the bottom-up input (connection with X area) to generate concrete and intuitive concepts, e.g. object type and location; some of them may merely use the top-down input (connection with Z area) to generate abstract concepts, and the rest of them may use information from both inputs. Thus, the Y area is gradually developed into different regions with the specific connections for various concepts. In fact, early processing area and later processing area are found in human vision system by experiments of neuroscience. However, the principle and mechanism of vision cortex division is still a significant puzzle. So far, there hasn't been a reasonable model in neuroscience or Artificial Intelligence which can solve this problem.

In this paper, a new biologically inspired mechanism — cross-domain synaptic maintenance will be proposed and introduced in our new WWN-9, which could be one of the mechanisms for emergence of early area and later area in Y in a perspective of computer simulation. In WWN-9, each neuron in Y area has three domains of input: bottom-up $\mathbf{b}$, lateral $\mathbf{l}$ and top-down $\mathbf{t}$. In the complex tasks such as recognition of object groups, the differentiation of neurons in Y means that some domains of irrelative input should be trimmed completely. The cut and retention of synapses from a domain depends on the correlations of weights and inputs which are indicated by ratio of standard deviation to the expected deviation among all the domains. We believe that synaptic maintenance works among all the three input domains (bottom up, lateral and top down) and combines the second order statistical characteristics of these inputs to adjust the contribution of three domains to Y area.

The existing versions of WWNs which recognize single object in natural backgrounds only have intuitive and independent motors such as type, position and scale. In order to study the internal connection of motor area, we teach the network the concept of object groups based on single objects that have been learnt previously. In real world most objects are composed of several components which have different kinds of relationships and several objects gather in pairs or as a group. For example, relationship of two animals involves friendly relationship, mother-child relationship, enemy relationship and so on. In the wild, to recognize the relationships illustrated in Fig. 1 is an essential skill for animals to survive. The relations between components of an object or members of a group are important concepts for a human or robot baby to understand. The grouping of objects which we will focus on is a special relationship between two objects. How to detect and recognize object groups in natural backgrounds is still challenging. The network should have ability to learn the concept of group, the position relation of two objects. The concept of object group is based on type and position of each single object. As the increasingly complexity of motor areas, the neurons in Y need to differentiate adaptively to play a specific role between different parts of sections. By cross-domain synaptic maintenance, the Y neurons can dynamically cut some synapses connecting with irrelevant input domains or sub domains.

In the remainder of the paper, Section II overviews the architecture and operation of WWN-9. Section III presents some important concepts and algorithms in the network. Experimental results are reported in Section IV. Section V gives the concluding remarks.

II. Network Overview

Similar to WWN-6, the architecture of WWN-9 consists of three areas, X area (sensory ends / sensors), Y area (internal brain inside the skull) and Z area (motor ends / effectors) (shown in Fig. 2).

X acts as the retina, which perceives the inputs and sends signals to internal brain Y. Z serves as both input and output. When the environment supervises Z, Z is the input to the network. Otherwise, Z gives the output to drive effectors which act on the real world. Functionally, Z is used as the hub for emergent concepts (e.g., location, scale and type), abstraction (many forms mapped to one equivalent state), and reasoning (as goal-dependant emergent action). In our paradigm of WWN-9, besides two same categories of concepts in WWN-6, the location and the type of the foreground object which corresponds to Location Motor (LM) and Type Motor (TM) respectively, a new category of concept Group Motor (GM) is added, which corresponds to group or pair of multi-objects in more abstract level. It includes the concepts of object pairs, such as AB, AC, CD, which are location invariant. Internal brain Y is a "bridge" linking both X and Z as its two "banks" through 2-way connections. It is worth to note that Y is completely inside the closed skull, which is off limit to the teachers in the external environments.

Different with WWN-6, the morphology of WWN-9 structure is expected to be changed with training automatically. With the training of the network, neurons in Y gradually differentiate into two types, early processing neurons and later processing neurons. The former mainly handle inputs from X while the latter mainly process signals from Z (TM, LM and GM). Thus the network will autonomously develop from the initial stage (shown in Fig. 2) to the mature stage (shown in Fig. 3).

III. Concepts and Theory

A. Receptive fields perceived by Y neurons

Neurons in Y have earlier connections with both X and Z in initial stage. A part of neurons in Y whose synapses with X are arranged regularly and tightly in space have fixed and intensive receptive fields, thus they can detect stable features from input image. These neurons in Y have the local receptive fields from the retina. Suppose the receptive field is $a \times a$, the neuron (i, j) perceives the region $R(x, y)$ in the input image ($i \le x \le (i+a-1)$, $j \le y \le (j+a-1)$), where the coordinate (i, j) represents the location of the neuron on the two-dimensional plane and similarly the coordinate (x, y) denotes the location of the pixel on the input image.

There are also some neurons in Y whose synapses with X are arranged loosely and randomly in space have big but sparse receptive fields, thus their input are generally irrelevant to be used as stable features for recognition. Due to the mechanism of cross-domain synaptic maintenance, the latter processing neurons will gradually cut the synapses connecting with X and process signals from motor areas specifically.

B. Pre-response of the Neurons

It is desirable that each brain area uses the same area function f, which can develop area specific representation and generate area specific responses. Each area A has a weight vector $\mathbf{v} = (\mathbf{v}_b, \mathbf{v}_t)$. Its pre-response value is:

$$r(\mathbf{v}_b, \mathbf{b}, \mathbf{v}_t, \mathbf{t}) = \dot{\mathbf{v}} \cdot \dot{\mathbf{p}} \tag{1}$$

where $\dot{\mathbf{v}}$ is the unit vector of the normalized synaptic vector $\mathbf{v} = (\dot{\mathbf{v}}_b, \dot{\mathbf{v}}_t)$, and $\dot{\mathbf{p}}$ is the unit vector of the normalized input vector $\mathbf{p} = (\dot{\mathbf{b}}, \dot{\mathbf{t}})$. The inner product measures the degree of match between these two directions of $\dot{\mathbf{v}}$ and $\dot{\mathbf{p}}$, because $r(\mathbf{v}_b, \mathbf{b}, \mathbf{v}_t, \mathbf{t}) = \cos(\theta)$ where θ is the angle between two unit vectors $\dot{\mathbf{v}}$ and $\dot{\mathbf{p}}$. This enables a match between two vectors of different magnitudes. The pre-response value ranges in $[-1, 1]$.

In other words, if regarding the synaptic weight vector as the object feature stored in the neuron, the pre-response measures the similarity between the input signal and the object feature.

The firing of a neuron is determined by the response intensity measured by the pre-response (shown as Equation 1). That is to say, If a neuron becomes a winner through the top-k competition of response intensity, this neuron will fire while all the other neurons are set to zero. In the network training, both motors' firing is imposed by the external teacher. In testing, the network operates in the free-viewing mode if neither is imposed, and in the location-goal mode if LM's firing is imposed, and in the type-goal mode if TM's is imposed. The firing of Y (internal brain) neurons is always autonomous, which is determined only by the competition among them.

C. Top-k Competition

Top-k competition takes place among the neurons in the same area, imitating the lateral inhibition which effectively suppresses the weakly matched neurons (measured by the pre-responses). Top-k competition guarantees that different neurons detect different features. The response r'_q after top-k competition is

$$r'_q = \begin{cases} (r_q - r_{k+1})/(r_1 - r_{k+1}) & \text{if } 1 \le q \le k \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

where r_1, r_q and r_{k+1} denote the first, qth and $(k+1)$th neuron's pre-response respectively after being sorted in descending order. This means that only the top-k responding neurons can fire while all the other neurons are set to zero.

D. Hebbian-like learning of neurons

The concept of neuronal age will be described before introducing Hebbian-like learning. Neuronal age represents the firing times of a neuron, i.e., the age of a neuron increases by one every time it fires. Once a neuron fires, it will implement Hebbian-like learning and then update its synaptic weights and age. A neuron with lower age has higher learning

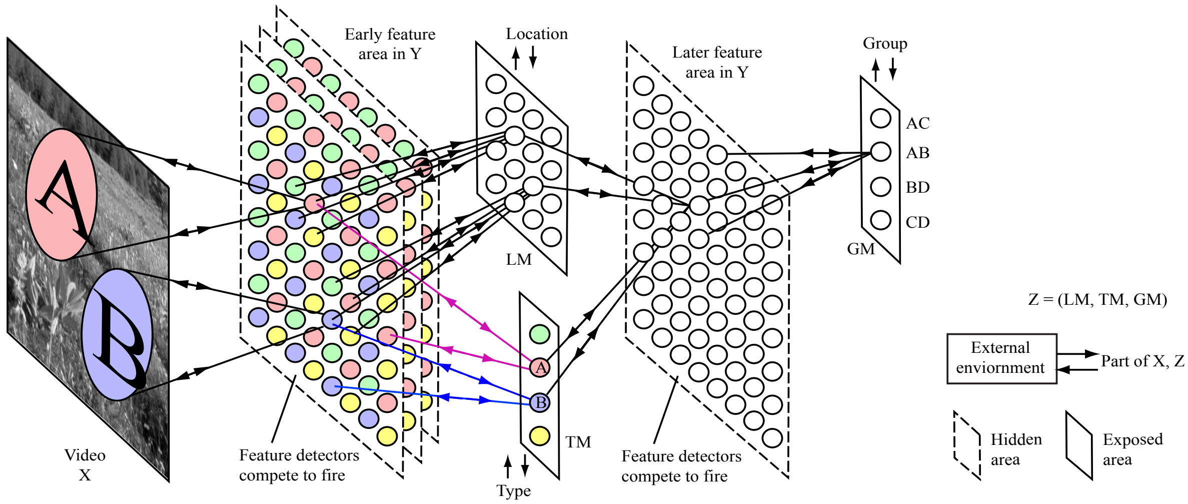


Fig. 3. The architecture of WWN-9 in mature stage through development using cross-domain synaptic maintenance. One two-way arrow means two one-way arrows because of the limited space.

rate and lower retention rate. Just like human, people usually lose some memory capacity as they get older. At the "birth" time, the age of all the neurons is initialized to 1, indicating 100% learning rate and 0% retention rate.

Hebbian-like learning is described as:

$$\mathbf{v}_j(n) = w_1(n)\mathbf{v}_j(n-1) + w_2(n)r'_j(t)\dot{\mathbf{p}}_j(t) \tag{3}$$

where $r'_j(t)$ is the response of the neuron j after top-k competition, n is the age of the neuron (related to the firing times of the neuron), $\mathbf{v}_j(n)$ is the synaptic weights vector of the neuron, $\dot{\mathbf{p}}_j(t)$ is the input patch perceived by the neuron, w_1 and w_2 are two parameters representing retention rate and learning rate with $w_1 + w_2 \equiv 1$. These two parameters are defined as following:

$$w_1(n) = 1 - w_2(n), \qquad w_2(n) = \frac{1+u(n)}{n} \tag{4}$$

where $u(n)$ is the amnesic function:

$$u(n) = \begin{cases} 0 & \text{if } n \le t_1 \\ c(n-t_1)/(t_2-t_1) & \text{if } t_1 < n \le t_2 \\ c + (n-t_2)/r & \text{if } t_2 < n \end{cases} \tag{5}$$

where $t_1 = 20, t_2 = 200, c = 2, r = 10000$ [7].

E. Synaptic maintenance

Synaptic maintenance seems to be conducted by every neuron in the brain. Each neuron, generated from neurogenesis (mitosis) autonomously decides where in the network to connect. In the Developmental Network (DN), each neuron does not have a pre-selected feature to detect. The role of each neuron is dynamically determined through its interactions with other neurons — known as the process of autonomous development.

Suppose that a neuron has an initial input vector $\mathbf{p}$ defined by all its spines where synapses sit. It would like to remove all the synapse components in $\mathbf{p}$ that is irrelevant to its post-synaptic firing (i.e., cluttered backgrounds in vision), while minimizing the removal of those relevant components (i.e., a foreground object). The removal is based on statistical score of match, between the pre-synaptic activities and the synaptic conductance (weight).

The known synaptic factors includes acetylcholine, agrin, astrocytes, neuroligins, SynCAM and Clq. Lichtman and co-workers [3] showed that partial blockage of the acetylcholine receptor (AChR) leads to retraction of corresponding presynaptic terminals. We believe that ACh signals expected uncertainty, or "this neuron predicts this pre-synaptic line pretty well."

Suppose that the input to a neuron is $\mathbf{p} = (p_1, p_2, ..., p_d)$ and its synaptic weight vector is $\mathbf{v} = (v_1, v_2, ..., v_d)$. Since each synapse sits on its spine, we should consider that this synaptic weight vector to be the composite effect of both the spines and the synapses.

Acetylcholine (ACh) originates from the basal forebrain, indicating expected uncertainty. We model how to neuromorphically measure expected uncertainty. When top-k neurons fire with value y, its synapse indicates the mean of the pre-synaptic activities x_i.

$$v_i = E[yp_i \mid \text{the neuron fires}] \tag{6}$$

using amnesic average. The standard deviation of match between v_i and p_i is a measure of expected uncertainty for each synapse i:

$$\sigma_i = E[|v_i - p_i| \mid \text{the neuron fires}] \tag{7}$$

is the expected uncertainty for each synapse, modeled by ACh. Mathematically, σ_i is the expected standard deviation of the match by the synapse i.

However, it must start with a constant value and wait till all the weights of the neuron have good estimates of w_i. Suppose that $\sigma_i(n)$ is σ_i at *firing age* n. Every synapse starts with the standard deviation of uniform distribution in $[-\delta, \delta]$, when $n \le n_0$. Then, the synapse i starts with normal incremental average. Finally, we use a constant asymptotic learning rate to enable the standard deviation to continuously to be plastic. The expression is as follows:

$$\sigma_i(n) = \begin{cases} 1/\sqrt{12} & \text{if } n \le n_0 \\ w_1(n-n_0)\sigma_i(n)+ & \\ w_2(n-n_0)|v_i - p_i| & \text{otherwise} \end{cases} \quad (8)$$

where $w_1(n)$ and $w_2(n)$ are defined as the same as the EQ. (4) and we set the latency for the synaptic maintenance $n_0 = 10$, to wait synapse weights (the first order statistics) to get good estimates first through the first n_0 updates before the standard deviation σ_i (the second order statistics) can have reasonable observations. The default estimate for σ_i, $1/\sqrt{12}$, is needed at early ages. Note that when calculate the learning rate and retention rate, the age of neurons should subtract n_0 to ensure the synapse value begin to update at a learning rate of 100%.

Each neuron should dynamically determine which synapse should keep active and which synapse should be retracted depending the goodness of match.

The expected goodness of match is indicated by the expected uncertainty, which involves a type of neuromodulators called Acetylcholine (ACh). The expected synaptic deviation among all the synapses of a neuron is defined by:

$$\bar{\sigma}(n) = \frac{1}{d(n)} \sum_{i=1}^{d(n)} \sigma_i(n) \quad (9)$$

where we assume that three input domains of input $\mathbf{b}, \mathbf{l}, \mathbf{t}$ are considered as one integrated source of input.

F. Cross-domain synaptic maintenance

In the DN model, each neuron in Y has three domains of input: bottom-up $\mathbf{b}$, lateral $\mathbf{l}$ and top-down $\mathbf{t}$. Some domain has several sub-domains. For example, the top-down domain $\mathbf{t}$ is Z area which has three sub-domains, LM, TM and GM. The bottom-up domain $\mathbf{b}$ is X area and lateral domain $\mathbf{l}$ is Y area.

Note that each pre-synaptic activity has already normalized into the range $[0, 1]$ where 0 means not firing and 1 means firing (or brightest pixel). Therefore, the σ_{ij} from the i-th domain and j-th neuron can be compared.

However, since the dimension is very different across different domains, we need to make sure that a low-dimensional domain plays a considerable role as a high-dimensional domain. Thus, the expected synaptic deviation in Eq. (9) should be modified to the following multi-domain version:

$$\bar{\sigma}(n) = \frac{1}{c(n)} \sum_{i=1}^{s(n)} \left(\frac{\beta_i}{d_i(0)} \sum_{j=1}^{d_i(n)} \sigma_{ij}(n) \right) \quad (10)$$

where $s(n)$ is the number of domains including 1-D domain, d_i the dimension of domain i, β_i the percentage of energy for domain i, $d(n)$ the current dimension of the input source $\mathbf{p}$ after one or more domains have probably been cut, and $c(n)$ is to make sure that the sum of all weights is one:

$$c(n) = \sum_{i=1}^{s(n)} \frac{\beta_i}{d_i(0)} d_i(n) \quad (11)$$

Note that $d_i(0)$ and $\beta_i(0)$ are used to set initial weights of each domain but this expression allows one or more domains (or sub-domains) to be cut completely. For example, the bottom-up, lateral, Z domains have energy percentages 1/3, 1/3 and 1/3, respectively, to sum to 100%.

Let

$$w_i(n) = \frac{\beta_i}{c(n) d_i(0)} \quad (12)$$

We have the expected synaptic deviation:

$$\bar{\sigma}(n) = \sum_{i=1}^{s(n)} w_i(n) \left(\sum_{j=1}^{d_i(n)} \sigma_{ij}(n) \right) \quad (13)$$

This means that in a domain with fewer synapses each synapse has more voice in vote for the $\bar{\sigma}(n)$.

We still define the neuronal samples of relative ratios as novelty transmitters Norepinephrine (NE) :

$$r_{ij}(n) = \frac{\sigma_{ij}(n)}{\bar{\sigma}(n)} \quad (14)$$

The cross-domain synaptic factor that uses three linear segments is

$$f(\sigma_{ij}, \bar{\sigma}) = \begin{cases} 1 & \text{if } \sigma_{ij}/\bar{\sigma} < \beta_s \\ (\beta_b - r_{ij})/(\beta_b - \beta_s) & \text{if } \beta_s \le \sigma_{ij}/\bar{\sigma} \le \beta_b \\ 0 & \text{otherwise} \end{cases} \quad (15)$$

We would like to retract synapse whose $\sigma(n)$ is relatively large. When a domain i has its relative ratio of deviation all higher than β_b, this domain has all its synapses cut off. From the (domain weighted) distribution of $\{r_{ij}\}$ in different domains, we get separate β_s and β_b for synapse connecting with X and Z.

G. Synapse trimming

Trimming can be considered as the maintenance of spine-synapse combination.

We would like to define the trimming of weights vector $\mathbf{v} = (v_1, v_2, ..., v_d)$ to be

$$v_i \leftarrow f_i v_i \quad (16)$$

$i = 1, 2, ..., d$.

Similarly, trim the input vector $\mathbf{p} = (p_1, p_2, ..., p_d)$ where $\mathbf{p} = (\mathbf{b}, \mathbf{l}, \mathbf{t})$.

Then the calculation of trimmed response should be modified accordingly. Therefore the synapse factor dynamically determines whether the corresponding synapse provides a full supply, no supply, or in between.

H. Differentiation of neurons in Y area

In our model, some neurons in Y cut off most of their synapses connecting with X and develop into the later area in Y. They receive the signals only from Z and combine information from various motors (e.g., TM and LM) for abstract reasoning and concluding. On the other hand, Some Y neurons cut off their synapses connecting with GM and develop into the early area in Y. By the mechanism of cross-domain synaptic maintenance, the network develops from

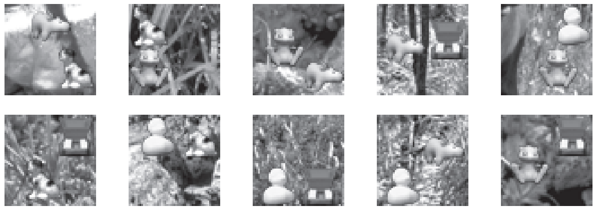


Fig. 4. The object group samples consisted of two single objects with the size of 22×22 in the background with the size of 51×51. The single object data set includes cat, dog, truck, pig and duck.

its initial stage to its mature stage with the emergence of early area and later area in Y. We will describe the details of developing process theoretically.

The neurons near the sensor X have continuous receptive field as big as the foreground objects. Through Hebbian learning, they can learn features of objects. Suppose one neuron has learnt feature of object A and connected with type A in TM and position of A in LM. However, it connects with more than one group in GM (e.g., AB, AC, AD). A neuron whose age exceeds the latency for the synaptic maintenance (i.e., n_0) begins to do the synaptic maintenance. When it fires, the inputs from TM, LM and X are always the same with related weights while the inputs from GM are different from the weights in Y. So the standard deviations of inputs from GM are considerable while that of the inputs from TM, LM and X are zeros. According to the EQ. (10), the standard deviations of inputs from GM are larger than $\bar{\sigma}$ and these synapses will be cut off.

For the neurons which are far from the sensor X and near the effector Z, each neuron learns a group of objects in possible locations. So one neuron learns a group (e.g., AB) connects with two types in TM (e.g., A and B) two locations in LM and a group in GM. Under supervised learning, the supervision given by teachers is certainly correct, so the deviations of inputs from TM, LM and GM are nearly zeros. Conversely, the receptive field of neurons far from X is not continuous and somewhat random, and most of its inputs from X are the pixels from backgrounds. Because standard deviations of match between background pixels and synaptic weights are relatively large, most of the synapses connecting with X will be cut and the synapses connecting with TM, LM and GM will be retained. Therefore these neurons develop into later area in Y.

In a word, the cross-domain synaptic maintenance adjusts the role of neurons in internal brain based on their initial location and distribution, and refines their earlier coarse connections to the sensor X and the motor Z. The synapses of neurons are dynamically self-developing all the time rather than static after birth.

IV. Experiments and results

A. Experiment Design

In our experiment, all the background patches are selected from 13 natural images and cropped into 51×51 pixels. The foreground objects are selected from the MSU 25-objects data set. Totally 5 different types of objects are paired into 10 groups. The possible locations of object are 30×30 (i.e., LM has 30×30 neurons). Assume that the single objects are A, B, C, D and E, then all the combinations of two-object pair are AB, AC, AD, AE, BC, BD, BE, CD, CE and DE. There is no limitation on the member positions in the same group except their distance can be regarded as a group. The input images for training and testing are shown in Fig. 4.

At each epoch, the network learns single objects and object groups successively with supervision of an external teacher. During the training of single objects in the group, the teacher only need provide the corresponding correct information in TM and LM so that no neuron in GM fires. While during the training of object groups, the teacher should provide not only the pair label of the group, but also the types and the locations of two members in the group.

For single object training, each single object needs to be trained at every possible position. However, for object group training, it is unnecessary to do such exhaustive training at all the position combinations since the later area is location invariant. A test in the free-viewing mode is performed after training at each epoch. The free-viewing mode means that the network works without any teacher's supervision of type or location.

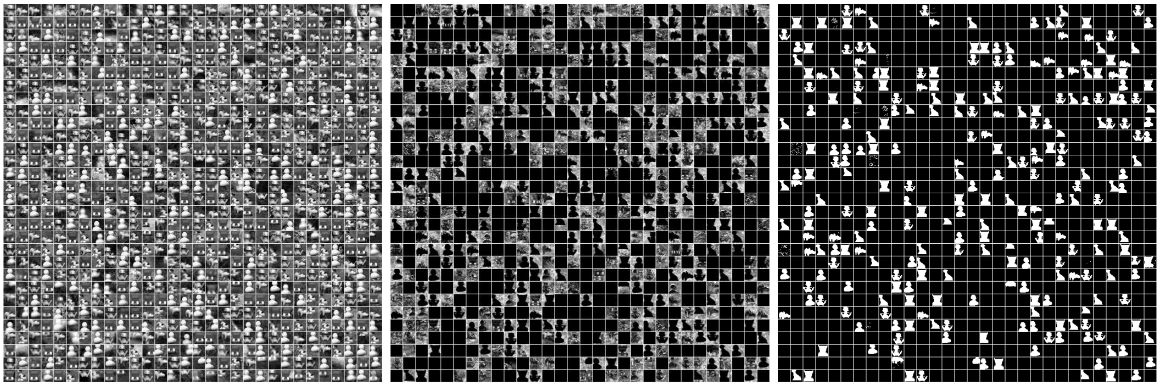
B. Visualization of synapse weights

To observe the change of synapse weights during the training period and understand the details of cross-domain synaptic maintenance well, the synaptic weights of neurons in Y are visualized in images consisting of a grid of small square patch. The weights of synapse could be observed by the color or intensity of pixels in each image patch.

The bottom-up weights and their standard deviations, and the bottom up synapse factors of the neurons in the early area of Y are displayed in Fig. 5. The standard deviations of synapse weights in TM and LM domain are nearly zeros, because the type and location of foreground object provided by the teacher during training are absolutely right. The bottom-up synapse weights corresponding to the background pixels have relatively large standard deviations and will be trimmed at different extent. From the Fig. 5(c), it is obvious that the contours of the foreground objects can be outlined automatically via synaptic maintenance. The black image patch indicates that the age of corresponding neuron is not large enough to do the synaptic maintenance.

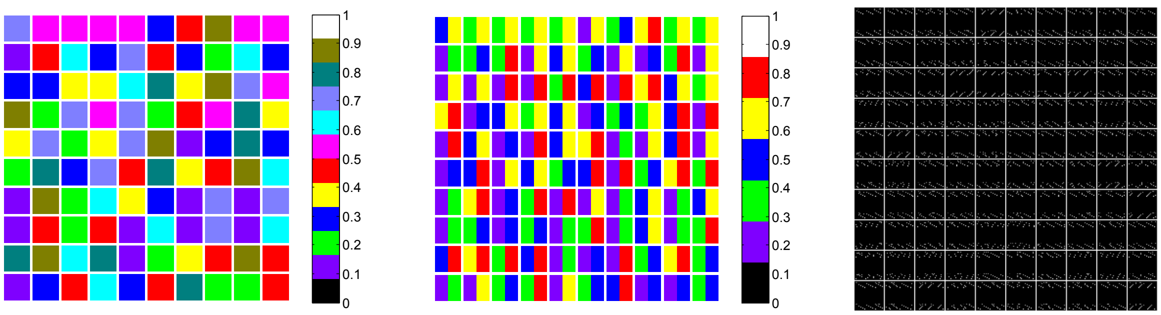
The top-down synapse weights of neurons in later area of Y are shown in Fig. 6. We can see that each neuron in later area learns only one type of group (Fig. 6(a)) and its corresponding types of two members (Fig. 6(b)). We also found that there are more than one group position (i.e., a pair of highlighted single object positions) in each small image patch (Fig. 6(c)) , which shows that each neuron in later area can do the group recognition with location invariance.

For neurons in later area of Y, their inputs from X are not stable statistically, thus the synapses connecting from X will be cut off as the training time increases, as shown in Fig. 7. Because of the random connection from X to later



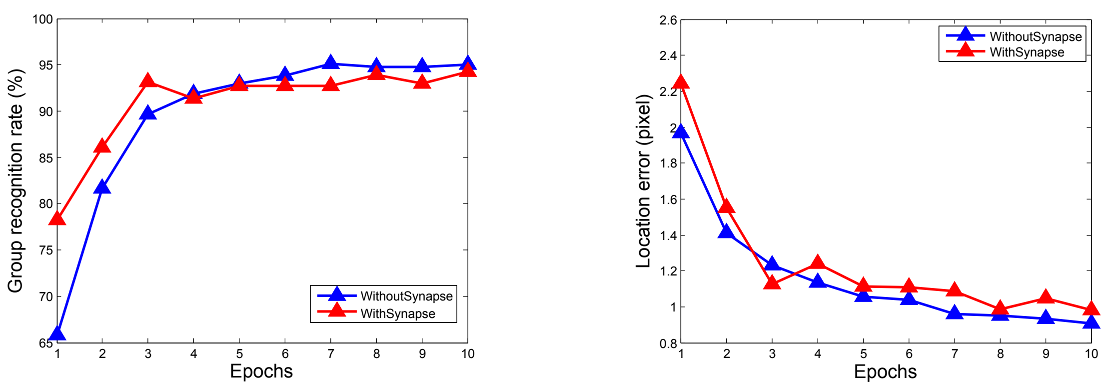
(a) Bottom-up weights (b) Standard deviations of bottom-up weights (c) Bottom-up synapse factors

Fig. 5. Visualization of the bottom-up weights of early area neurons in Y (only one of the three depths shown here). Each small square patch in (a) visualized the bottom-up weight vector of one neuron, i.e., the feature of one training object. Each small patch in (b) visualized the standard deviations of bottom-up weights updated by Hebbian learning. Each small patch in (c) visualized the bottom-up synapse factor (vector) of one neuron. The white pixels in the image patch correspond to the foreground pixels (object) and the black pixels correspond to the background pixels.



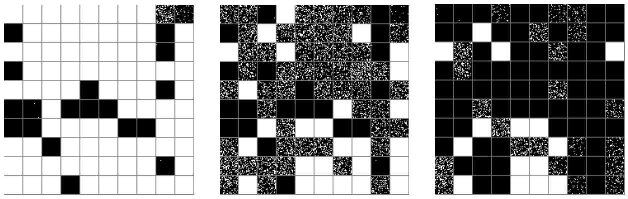
(a) GM (b) TM (c) LM

Fig. 6. Visualization of the top-down weights of later area neurons in Y. Each small square patch corresponds to one neuron, which refers to the different type of top-down weights: weights from GM (a), weights from TM (b) and weights from LM (c). Block color in (a) represents the type of the specific group, and the color bar at the right side is the color map correspond to the 10 training groups. Block color in (b) represents the type of the specific single object. Each image patch (30×30) in (c) presents top-down weights from LM with the single object positions in the group highlighted by the white or gray pixels.



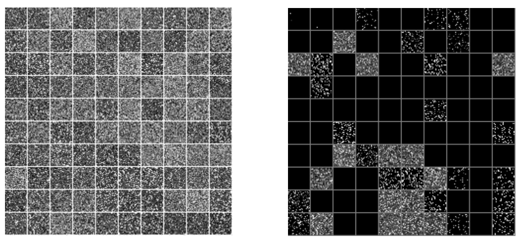
(a) Group Recognition Rate (b) Location Error

Fig. 9. Network performance variation in 10 epochs with/without synaptic maintenance for 10 groups.



(a) Epoch 3 (b) Epoch 7 (c) Epoch 10

Fig. 7. Visualization of the bottom-up synapse factors of later area neurons in Y. The pixel values in each small image patch represent the retention rate of bottom-up synapses for one neuron in Y. The white pixels represent the retained synapses and the black ones represent the trimmed synapses. With the training going on, some of the synapses are cut off gradually.



(a) Untrimmed Weights (b) Trimmed weights

Fig. 8. Visualization of the bottom-up weights of later area neurons in Y. Each small square patch in (a) visualized a neuron's bottom-up weights vectors before trimming. Each small square patch in (b) visualized the weights of a neuron trimmed by synaptic maintenance.

area of Y, the bottom-up inputs change frequently and the synapse weights updated by Hebbian learning look like in a mess, as shown in Fig. 8(a). This maybe results from random selection of the synapse weights from X to later area of Y in each epoch during the training, which are always irrelevant with the inputs. Therefore, stable weights can't be available and the standard deviations of bottom-up synapse weights are large. Compared with steady supervision from top-down synapses, most of the bottom-up synapses will be trimmed, as shown in Fig. 8(a).

C. *Comparison of WWN with/without cross-domain synaptic maintenance*

In this experiment, the network is training with insufficient resources. Here, "sufficient resource" means that each Y neuron only corresponds to one unique case, i.e., one type of object at one location. For example, in this experiment, the layer depth of Y is set to 3, therefore the network has sufficient resources to learn 3 foreground objects but to learn 5 foreground objects, it is short of $(5-3)/5 = 40\%$ resources.

The performance variation of the network in 10 epochs with/without cross-domain synaptic maintenance is shown as Fig. 9. One epoch of training means that the network learns all the objects with all the possible locations in the images and each image is used for three iterations in one epoch.

The recognition rate of group is the percentage of the correct answers among all the testing cases. The location error is the distance between true position and predicted position of the two members in a group. The performance of group recognition improves with the increase of epochs. With cross-domain synaptic maintenance, the recognition rate reaches 95% and the location error is about 1 pixel in 10th epoch, nearly as well as that without cross-domain synaptic maintenance. Theoretically, the cross-domain synaptic maintenance should improve the performance in group recognition because it can prevent the influence from the irrelevant inputs and lead to more accurate clustering. A possible reason for current unsatisfying result is that for late area of Y, the input from X is already much less than the inputs from the other domains even without trimming so that the synaptic maintenance takes no effect. Thus, the mechanism and application potential of cross-domain synaptic maintenance will be further studied in details.

From the visualization of synapse weights, we can see that cross-domain synaptic maintenance plays an important role in the emergence of early area and later area in Y. Although the recognition rate and location error are not improved, it could avoid most of response computing of the irrelevant inputs. As the large computing amount in computer simulation of neural networks, this mechanism could reduce the computational time and help the network to work real time. Besides, the reduction of synapse computing amount can lower energy consumption which is of great importance for human and animals.

V. Conclusions and Future Work

The biologically inspired mechanism of cross-domain synaptic maintenance seems to improve the performance of multi-task learning — learning individual objects and learning object groups. The cross-domain synaptic maintenance might be useful for the emergence of early areas and later areas in the brain. In the future work, object groups in natural videos will be directly used in experiments. An ongoing work is to apply cross-domain synaptic maintenance to lateral connections.

References

[1] G. A. Carpenter, S. Grossberg, and J. H. Reynolds. ARTMAP: Supervised real-time learning and classification of nonstationary data by a self-organizing neural networks. *Neural Networks*, 4:565–588, 1991.

[2] G. E. Hinton, S. Osindero, and Y-. W. Teh. A fast learning algorithm for deep belief nets. *Neural Computation*, 18:1527–1554, 2006.

[3] C. M. McCann, Q. T. Nguyen, H. S. Neto, and J. W. Lichtman. Rapid synapse elimination after postsynaptic protein synthesis inhibition in vivo. *Journal Physiol.*, 27(22):6064–6067, 2007.

[4] T. Serre, L. Wolf, S. Bileschi, M. Riesenhuber, and T. Poggio. Robust object recognition with cortex-like mechanisms. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 29(3):411–426, 2007.

[5] Yuekai Wang, Xiaofeng Wu, and Juyang Weng. Synapse maintenance in the where-what networks. In *Neural Networks (IJCNN), The 2011 International Joint Conference on*, pages 2822–2829. IEEE, 2011.

[6] J. Weng. Why have we passed "neural networks do not abstract well"? *Natural Intelligence: the INNS Magazine*, 1(1):13–22, 2011.

[7] J. Weng and M. Luciw. Dually optimal neuronal layers: Lobe component analysis. *IEEE Trans. Autonomous Mental Development*, 1(1):68–85, 2009.